ELO iSearch Leitfaden

[Stand: 27.09.2019 | Programmversion: 12.00.000]

Dieser Leitfaden soll sowohl den fachlichen Benutzer bei der Verwendung und beim Einsatz der vielfältigen Suchfunktionen unterstützen als auch den Administratoren wertvolle Hinweise bei der Installation und dem Betrieb der ELO iSearch bieten. Ziel ist es, die Leistungsfähigkeit der ELO iSearch im Arbeitsalltag auszuschöpfen und die sich bietenden Möglichkeiten optimal einzusetzen.

Inhalt

1 Finden statt suchen 4
2 Was leistet die ELO iSearch? 5
3 Die ELO iSearch in der Praxis 6
3.1 Grundlagen und die „2-Wege-Strategie“ 6
3.1.1 ELO iSearch 6
3.1.2 Verschlagwortungssuche 6
3.2 Volltextsuche 6
3.2.1 Linguistik 7
3.2.2 Sprachen 8
3.2.3 Mehrwortsuche 9
3.2.4 Phrasensuche 9
3.2.5 Thesaurus 9
3.3 Suchbereich/Durchsuchte Felder 10
3.3.1 Volltext 10
3.3.2 Kurzbezeichnung 11
3.3.3 Indexfelder 11
3.3.4 Zusatztext 11
3.3.5 Feed 11
3.4 Filter 11
3.4.1 Eintragstyp 12

3.4.2 Datum/Ablagedatum ... 13
3.4.3 Abgelegt von/Bearbeiter ... 17
3.4.4 Maske ... 17
3.4.5 Alle Indexfelder ... 18
3.4.6 Indexfeld auswählen ... 18
3.4.7 Numerische Indexfelder ... 21
3.4.8 Feed ... 21
3.4.9 Dateigröße ... 22
3.4.10 Suche im aktuellen Ordner ... 23
3.5 Suche speichern ... 23
3.5.1 Standardfilter ... 24
3.5.2 Suchfavoriten ... 25
3.5.3 Suchfilter administrieren ... 26
3.6 Trefferliste ... 27
3.6.1 Treffermenge und -seiten ... 28
3.6.2 Darstellung als Liste, Kachel oder Baum ... 29
4 Technische Beschreibung ... 32
4.1 Architektur ... 32
4.1.1 Elasticsearch ... 32
4.1.2 Lucene ... 32
4.1.3 Sprachunterstützung ... 33
4.1.4 Verarbeitung mit ELO ... 33
4.1.5 Sicherheit ... 34
4.2 Installation und Upgrade ... 34
4.2.1 ELO Server Setup ... 34
4.2.2 Prüfungen nach der Installation ... 35
4.2.3 Volltextindexaufbau nach Upgrade von älteren ELO Versionen ... 36
4.2.4 Volltextindexaufbau im Hintergrund ... 38
4.3 Sizing und Performance ... 39
4.3.1 Allgemeine Aussagen zum Sizing ... 39
4.3.2 Optimierung der Indexierung ... 40

4.3.3 Optimierung der Suchgeschwindigkeit ... 40
4.3.4 Beispielhafte Systemsizings ... 40
4.3.5 Laufzeiten beim Volltextindexaufbau ... 41
4.4 Betrieb ... 41
4.4.1 Log-Datei ... 41
4.4.2 Auswertungen ... 42
4.4.3 Backup-Möglichkeiten und Backup-Strategien ... 43
4.5 Tipps und Tricks ... 46
4.5.1 ELO iSearch-Konfiguration ... 46
4.5.2 Dienstekonfiguration der Elasticsearch ... 51
4.5.3 SSL-Kommunikation ... 51
4.5.4 Indexserver-Plugins ... 52
4.5.5 Festlegung der Shard-Anzahl pro Index ... 52
4.5.6 Festlegung der Replica-Anzahl pro Index ... 53
4.5.7 Tokenisierung ... 54
4.5.8 Einbinden eigener Thesauri ... 54
4.5.9 Cluster-Betrieb ... 55
4.6 Powersuche ... 56
4.6.1 Anwendung der Powersuche ... 56
4.6.2 Feldnamen ... 58
4.6.3 Beispiele für Abfragen ... 58
4.7 ELO Analytics ... 58
4.7.1 ELO Analytics-Komponentenübersicht ... 58
5 Lizenzbestimmungen bei der Verteilung von Serverprozessen ... 60
6 Glossar ... 62

1 Finden statt suchen

„Alles auf der Erde lässt sich finden, wenn man nur zu suchen sich nicht verdrießen lässt.“

Philemon (griechischer Dichter, 361-263v. Chr.)

Bei einem Anstieg unstrukturierter Daten in modernen Informationssystemen von mehr als 30% im Jahr stellt das Finden relevanter Informationen eine immer größere Herausforderung dar.

Die ELO iSearch mit der integrierten Suchtechnologie „Elasticsearch“ nimmt diese Herausforderung an.

Die moderne Suchtechnologie der ELO iSearch kombiniert die Mehrwerte einer ganzen Reihe von Suchmöglichkeiten, wie die Volltextsuche, Suche in den Metadaten und die direkte Suche in einer hierarchischen Ablagestruktur. So wird die Leistungsfähigkeit der jeweiligen IT-Infrastruktur jederzeit sinnvoll eingesetzt. Bei der Suche werden Ausfallsicherheit, Redundanz, Skalierbarkeit sowie Geschwindigkeit berücksichtigt.

Mit all diesen Faktoren, aber auch mit ihrer Fähigkeit zur Mehrsprachigkeit und der Berücksichtigung branchenspezifischer Begrifflichkeiten, setzt die ELO iSearch einen hohen Standard und öffnet ein völlig neues Kapitel beim Finden von unternehmensrelevanten Informationen.

2 Was leistet die ELO iSearch?

Die ELO ECM Suite 11 bildet die Basis für eine nutzenorientierte und effiziente Bereitstellung von Informationen: Relevante Daten werden auch bei größeren Mengen in kürzester Zeit gefunden. Komfortabel und schnell können die richtigen Treffer ausfindig gemacht werden.

Für die tägliche Suche nach Informationen und Dokumenten wird im Arbeitsalltag jede Menge Zeit verschwendet. Zeit, die am Ende für die Erledigung der Kernaufgaben fehlt. Damit ist die zeitintensive Informationsrecherche einer der größten Kostentreiber.

Hierbei stehen verschiedene Suchfunktionen zur Verfügung, sodass Sie zielgerichtet auf benötigte Dokumente sowie auf Informationen aus integrierten Drittanwendungen, wie ERP-, CRM- oder E-Mail-Systeme, zugreifen können. Die ELO iSearch unterstützt Ihre Suchanfrage beispielsweise durch die Autovervollständigung von Suchbegriffen während der Eingabe, durch die Korrektur möglicher Tippfehler und durch die Suche nach ähnlichen oder verwandten Begriffen.

Zudem schlägt die ELO iSearch beispielsweise bei der Eingabe des Suchbegriffes „Vertrag“ auch Synonyme wie „Vereinbarung“ oder „Abkommen“ vor. Und selbst wenn kein eindeutiger Suchbegriff definiert werden kann, führt die ELO iSearch zum Ergebnis. Es werden zum Beispiel Begriffe zur Auswahl vorgeschlagen, die im Zusammenhang mit dem eingegebenen Suchwort stehen. Intelligente Wörterbücher sind darüber hinaus in der Lage, zusammengesetzte Begriffe wie zum Beispiel „Versicherungsvertrag“ in seine Grundbestandteile zu zerlegen. Auch das Zurückführen von Suchbegriffen auf die jeweilige Grundform ist möglich. So müssen Sie sich nicht mehr um Pluralformen, gebeugte Wortformen oder um beliebig zusammengesetzte Begriffe bei der Formulierung Ihrer Suchanfrage kümmern. Dies erledigt die Suchtechnologie automatisch.

Bei Eingabe eines falsch geschriebenen Suchbegriffs, wie zum Beispiel „Vertrahg“, erscheint zudem eine Meldung: „Meinten Sie vielleicht Vertrag?“. Diese intelligenten Assistenzsysteme fangen bereits frühzeitig Eingabefehler ab; damit wird das Suchergebnis verbessert oder oft sogar erst grundsätzlich sichergestellt.

Bei der strukturierten Suche in den Metadaten sind die Suchanfragen häufig sehr präzise und gezielt. Die Suche nach einer Rechnungs-, Angebots- oder Vertragsnummer über das gesamte Archiv funktioniert schnell und zuverlässig.

Eine weitere Herausforderung besteht darin, häufig verwendete, immer wiederkehrende oder komplexe Suchanfragen mit zum Teil aufwendig zusammengestellten Filteroptionen zu speichern und bei Bedarf schnell und komfortabel aufrufen zu können. Die bereits mit der ELO ECM Suite 10 eingeführten Kacheln zum Speichern und Verwalten von Suchfavoriten sind hierfür die perfekte Lösung.

3 Die ELO iSearch in der Praxis

3.1 Grundlagen und die „2-Wege-Strategie“

3.1.1 ELO iSearch

Die ELO iSearch ermöglicht die Suche im Volltext, also dem Text innerhalb eines Dokuments und in den zusätzlichen Informationen wie der Verschlagwortung. Sie bietet dabei verschiedene Möglichkeiten, Suchbereiche miteinander zu kombinieren. Der Volltext wird dabei analysiert und mit Suchregeln und Suchmustern (Algorithmen) verarbeitet.

Parallel zur SQL-Datenbank wird automatisch ein spezieller Suchindex aufgebaut und aktualisiert. So werden schnelle Antwortzeiten auch bei komplizierten Suchanfragen ermöglicht. Allerdings bedeutet dies auch eine leichte Zeitverzögerung, bis neue Dokumente mit der ELO iSearch gefunden werden können. Üblicherweise sollte diese Verzögerung unter einer Minute liegen.

Die weiteren Kapitel beziehen sich auf die Benutzung der ELO iSearch im ELO Java Client.

3.1.2 Verschlagwortungssuche

Die Verschlagwortungssuche ist eine zweite Suchform für spezielle Anforderungen. Sie sucht immer exakt nach den eingegebenen Werten direkt in der SQL-Datenbank und unterliegt daher nicht der Zeitverzögerung.

Im Vergleich zur ELO iSearch kann es bei derselben Sucheingabe zu unterschiedlichen Ergebnissen kommen. Typischerweise wird dabei die ELO iSearch eher mehr Treffer finden als die Verschlagwortungssuche, da die technologische Umsetzung der sprachenspezifischen Suchregeln und Suchmuster zum Beispiel auch die Mehrzahlform eines Suchwortes findet.

Grundsätzlich ist die ELO iSearch der modernere, flexiblere leistungsfähigere Ansatz und die bevorzugte Suchvariante in ELO.

3.2 Volltextsuche

Der Volltext bezeichnet alle Textinhalte eines Dokuments. Bei einer Rechnung wären dies beispielsweise neben Namen, Anschrift und Adresse des Empfängers auch die des Erstellers, die Rechnungspositionen, Beträge und auch eventuell vorhandene Zahlungshinweise.

3.2.1 Linguistik

Die Linguistik (Sprachwissenschaft) befasst sich mit den Bedeutungen und Regeln der menschlichen Sprache. Auf Basis solcher Regeln verwendet die ELO iSearch automatisierte Methoden, um Texte so aufzubereiten, dass sie möglichst einfach und komfortabel gefunden werden können. Dadurch lassen sich mit einfachen Suchworten auch Dokumente finden, die inhaltlich passen. Es ist also nicht nötig, die exakte Schreibweise im Dokument zu kennen. Dies ist vor allem immer dann praktisch, wenn es unterschiedliche Schreibvarianten gibt. Nachfolgend wird in diesem Zusammenhang von Suchmustern oder Sprach- und Suchregeln gesprochen.

3.2.1.1 Kompositazerlegung

Zusammengesetzte Wörter werden als Komposita bezeichnet. Gerade die deutsche Sprache verwendet davon sehr viele, zum Beispiel „Eingangsrechnung“ oder auch der berühmte „Donaudampfschifffahrtskapitän“. Diese werden in ihre Einzelteile zerlegt, also z. B. „Eingang“ und „Rechnung“.

3.2.1.2 Stammformenbildung

Bei dieser Bearbeitung wird die Einzahl aus Mehrzahlformen gebildet oder der Infinitiv zu deklinierten Verben gefunden. Zu dem Wort „Rechnungen“ wird die Einzahl „Rechnung“ ergänzt, zu „bezahlt“ die Stammform „zahlen“. Umgekehrt, aus der Bildung von Mehrzahlformen aus dem Singular eines Wortes, ist jedoch nicht möglich! Dies wird u. a. in der nachfolgenden Tabelle verdeutlicht und sollte bei der Auswahl und der Schreibweise des Suchbegriffes berücksichtigt werden.



Beachten Sie: Kompositazerlegung und Stammformenbildung finden bei der Indexierung statt. Um dieses Feature zu nutzen, müssen Sie bei der Suche die Stammform eingeben. Verwenden Sie als Suchwort eine gebeugte Form, wird nach exakt dieser Form als einzelnes Wort gesucht. Beispiele hierzu finden Sie in der folgenden Tabelle.

Volltext	Suchwort	Treffer
Rechnungen	Rechnungen	Ja
Rechnungen	Rechnung	Ja
Rechnung	Rechnungen	Nein
Eingangsrechnungen	Eingangsrechnungen	Ja
Eingangsrechnungen	Rechnungen	Nein
Eingangsrechnungen	Rechnung	Ja
Rechnung	Rechnung	Ja

3.2.2 Sprachen

Grundsätzlich muss man beachten, dass alle Suchregeln abhängig von der verwendeten Sprache gelten. Im Englischen wird eine Mehrzahl in der Regel durch ein angehängtes -s gebildet (zum Beispiel „book“ und „books“), im Deutschen gibt es hingegen verschiedene Varianten und auch viele unregelmäßige Formen (aus einem „Buch“ werden mehrere „Bücher“).

Die ELO iSearch verwendet je nach Sprache unterschiedliche Regeln. Es stehen somit nicht in jeder Sprache Suchregeln und -muster im gleichen Umfang zur Verfügung. Die meisten Optimierungen der Suchregeln und -muster beziehen sich auf die deutsche Sprache. Die englische Sprache wird aber ebenfalls sehr gut unterstützt.

Um die Dokumente sinnvoll automatisch bearbeiten zu können, wird in einem ersten Schritt die Sprache des Volltextes anhand der verwendeten Wörter ermittelt. Diese wird dann zur Suchregel- und Suchmusterbearbeitung des Textes verwendet. Probleme kann es daher bei Dokumenten geben, die unterschiedliche Sprachen verwenden.

Die Indexfelder werden immer mit der eingestellten/ausgewählten Archivsprache bearbeitet.

3.2.3 Mehrwortsuche

Geben Sie mehrere Wörter in das Suchfeld ein, wird nach Dokumenten gesucht, in denen alle diese Wörter vorkommen. Sie schränken also die Suche mit jedem zusätzlichen Wort weiter ein und finden im Idealfall genau das passende Dokument.

Um mehrere alternative Bezeichnungen zu suchen, trennen Sie diese mit einem Komma, zum Beispiel „Dienstreise, Kundentermin“. Es genügt dann, wenn eines der Suchwörter im Dokument vorkommt.

Beachten Sie: Damit die Mehrwortsuche funktionieren kann, müssen alle gesuchten Wörter in einem Bereich vorhanden sein (siehe Kapitel *3.3 Suchbereich/Durchsuchte Felder*).

Eine Suche nach „Firmenfest Tischplan“ in allen Suchbereichen findet also Dokumente, bei denen die Wörter „Firmenfest“ und „Tischplan“ entweder beide im Volltext vorkommen, beide in der Verschlagwortung stehen oder beide im Feed genannt werden. Sie würden keine Dokumente finden, in denen die Wörter über mehrere Bereiche verteilt sind, z. B. das eine Wort im Volltext und das andere in der Verschlagwortung.

3.2.4 Phrasensuche

Es können nicht nur einzelne Wörter im Dokument gesucht werden, sondern auch längere Begriffe und ganze Sätze. Setzen Sie hierzu die zusammenhängenden Wörter in Anführungsstriche. Diese müssen dann genau in dieser Reihenfolge direkt hintereinander im Volltext vorkommen.

Mit einer Suche nach „Heinrich Müller“ finden Sie genau diesen Namen. Ohne die Anführungszeichen würden Sie auch Dokumente finden, in denen die Wörter weiter auseinanderstehen, zum Beispiel „Teilnehmer waren Heinrich Schulz und Dieter Müller“.

3.2.5 Thesaurus

In die ELO iSearch ist ein Thesaurus integriert, der Synonyme zu dem eingegebenen Suchwort auflistet. Diese erscheinen, sobald die Eingabe eines Wortes mit einem Leerzeichen beendet wird.

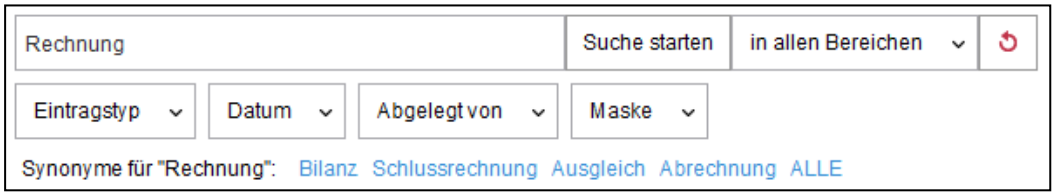


Abb. 1: Anzeige von Synonymen, in diesem Beispiel für „Rechnung“

Klicken Sie ein Synonym an, um es als Alternative in die Sucheingabe zu übernehmen. Der Thesaurus kann firmenspezifisch erweitert werden. Ebenso ist es möglich, Vorschläge zu entfernen. Lesen Sie dazu das Kapitel *4.5.9 Einbinden eigener Thesauri*.

3.3 Suchbereich/Durchsuchte Felder

Direkt neben dem Eingabefeld und der Schaltfläche „Suche starten" befindet sich die Auswahl des Suchbereichs. Hiermit können Sie einstellen, wo nach den eingegebenen Wörtern gesucht werden soll. Zur Auswahl stehen die drei Bereiche Volltext, Verschlagwortung und Feed. Der Bereich „Verschlagwortung" ist in „Kurzbezeichnung", „Indexfelder" und „Zusatztext" unterteilt. Die Standardeinstellung ist die Suche in allen Bereichen.

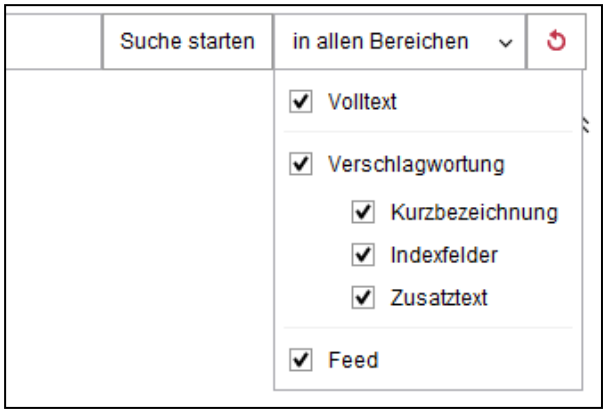


Abb. 2: Drop-down-Menü „Durchsuchte Felder"

Information: Es wird wie oben für die Volltextsuche beschrieben in allen Feldern mit den gleichen Suchregeln und -mustern gesucht. Stammformen werden erkannt oder Wortzusammensetzungen getrennt gesucht.

3.3.1 Volltext

Bei rein digitalen Dokumenten entspricht der Volltext immer genau dem Text im Dokument. Die Suche im Volltext kann also alle Inhalte des Dokuments finden. Im Gegensatz zu der Suche in Metadaten eignet sich die Volltextsuche dafür, nicht nur definierte Schlagworte oder Metainformationen zu finden, sondern alle Begriffe des Dokuments.

Gescannte Dokumente müssen erst mittels OCR in eine Textform umgewandelt werden. Dabei kann es immer zu Erkennungsfehlern einzelner Buchstaben oder ganzer Wörter kommen, welche bei der Suche das Finden des Dokuments verhindern. Zudem kann es durch Textblöcke und Spaltensatz zu einer Vermischung von Sätzen kommen, wodurch eine Phrasensuche (siehe Kapitel *3.2.4 Phrasensuche*) den Satz nicht finden kann.

3.3.2 Kurzbezeichnung

Dokumente und Ordner werden im Archiv mit ihrer Kurzbezeichnung angezeigt. Verwenden Sie daher eine möglichst treffende und korrekte Bezeichnung des Inhalts in der Kurzbezeichnung.

3.3.3 Indexfelder

Mit der Einstellung „Indexfelder" werden alle Indexfelder in allen Masken des Archivs in die Suche einbezogen. Suchen Sie in den Indexfeldern nach dem Wort „Müller", erhalten Sie beispielsweise folgende Treffer:

- „Müller Baustoffe GmbH" in einem Indexfeld „Lieferant"
- „Heinrich Müller" in einem Indexfeld „Kunde"
- Die „Müllerstraße 42" in einem Indexfeld „Anschrift"

3.3.4 Zusatztext

Der Zusatztext (eine Registerkarte in der Verschlagwortungsmaske) dient in ELO zur Aufnahme weiterer Informationen zu einem Dokument oder Ordner. Er bietet im Vergleich zu einem Indexfeld Platz für längere Texte und kann hier ebenfalls durchsucht werden.

3.3.5 Feed

Mit der Einstellung „Feed" werden alle Beiträge und Kommentare im Feed in Textform durchsucht.



Information: Im Feed sind auch die Versionsnummern und Versionskommentare enthalten, sowohl zu der aktuellen Arbeitsversionen eines Dokuments als auch zu älteren Versionen.

3.4 Filter

In vielen Fällen ist es sinnvoll, die reine Textsuche um strukturierte Informationen zur Einschränkung des Suchbereichs zu ergänzen. Die ELO iSearch bietet hierfür Filter an. Diese können über die Multifunktionsleiste eingeblendet werden und erscheinen dann als zusätzliche Schaltflächen unter dem Eingabefeld.

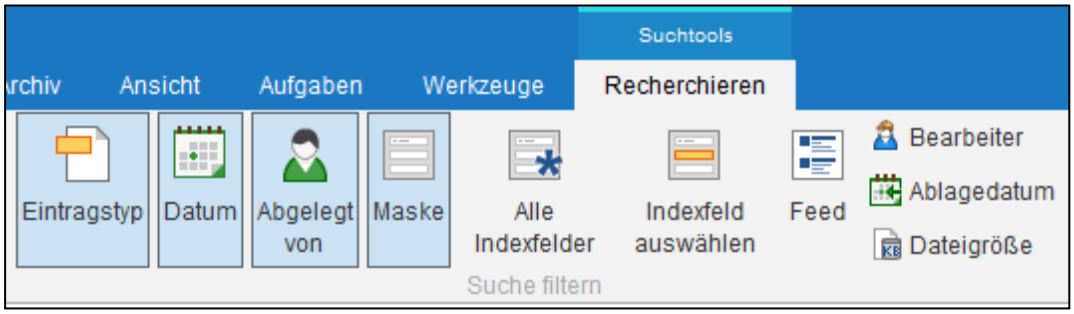


Abb. 3: Filter der ELO iSearch in der Multifunktionsleiste

Eine Reihe von Filtern wird standardmäßig unterhalb des Suchfeldes angezeigt. Es handelt sich dabei um die Filter, die am häufigsten verwendet werden. Welche Filter das sind, kann man sowohl global, als auch gruppenspezifisch und individuell einstellen, siehe Kapitel *3.5.2 Suchfavoriten*.

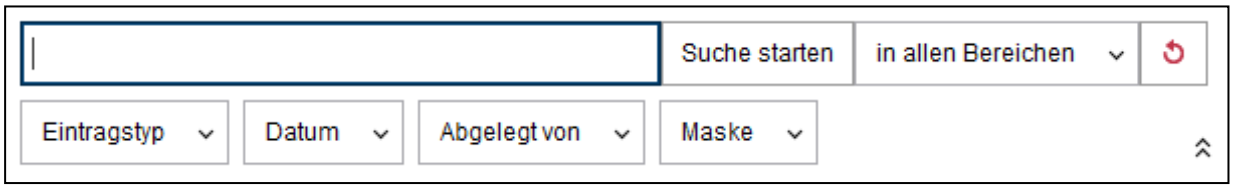


Abb. 4: Standardfilter unter dem Suchfeld

Durch Klick auf die Schaltfläche erscheint der jeweilige Filter und es können spezielle Suchangaben vorgenommen werden. Dabei ist fast immer eine Eingabe sowohl per Maus als auch per Tastatur möglich.

Die Filter berücksichtigen genauso wie die Suche die ELO Berechtigungsstruktur. Es werden also nur die Filterwerte zu den Einträgen angezeigt, ein Benutzer auch sehen darf.

3.4.1 Eintragstyp

Der Filter für den Eintragstyp zeigt, wie der Filter für die Maske, eine Liste der möglichen Werte an. Die Bedienung ist ebenfalls dieselbe, mit Search-as-you-type und Autovervollständigung oder Auswahl aus der Liste. Als Besonderheit werden hier ganz oben die zwei Sammelbegriffe „Dokumente“ und „Ordner“ angeboten, welche alle entsprechenden Eintragstypen zusammenfassen.

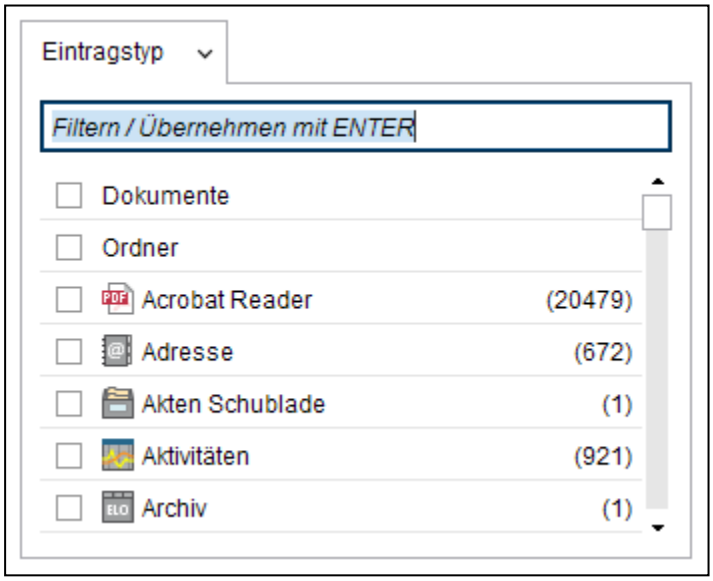


Abb. 5: Filter für den Eintragstyp

3.4.2 Datum/Ablagedatum

Der Filter für die Datumsangaben der Verschlagwortung bietet viele Möglichkeiten zur Suche nach einem Datum. Es wird dabei immer nach ganzen Tagen gesucht, also von 00:00:00 Uhr bis 23:59:59 eines Tages.

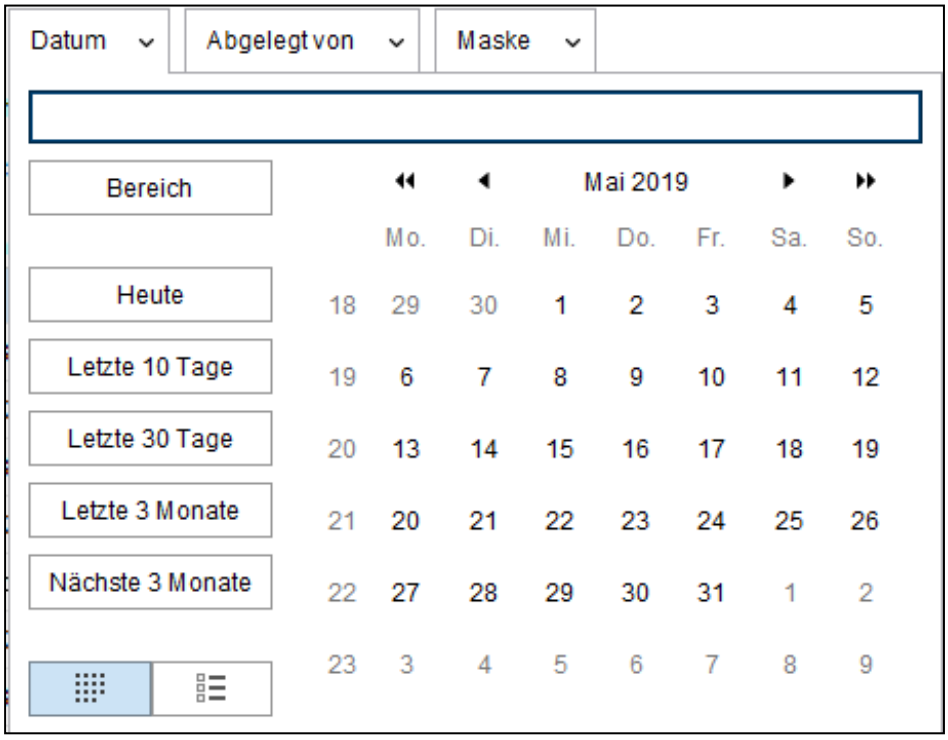


Abb. 6: Drop-down-Menü des Filters „Datum“

Ganz oben in der Abbildung 6 findet man ein Eingabefeld, in das man ein exaktes Datum eingeben kann. Hierbei verwendet der ELO Java Client die in der Konfiguration eingestellten Datumsformate. Einen Zeitraum gibt man mit drei Punkten als Trenner in der Mitte an, zum Beispiel 01.05.2019…10.05.2019. Die Eingabe müssen stets mit der ENTER-Taste abgeschlossen werden. Der Filter schließt sich und der Zeitraum wird in einem kleinen grauen Kasten dargestellt.



Abb. 7: Eingegebener Datumsfilter

Um den Zeitraum jedoch noch komfortabler einzugeben, kann man alternativ auf den Modus „Bereich" mit zwei Eingabefeldern wechseln. Nach der Eingabe des Wertes „von" springt der Fokus automatisch in das Feld „bis" und man kann beide Felder mit der Maus (Klick) über den Kalender befüllen.

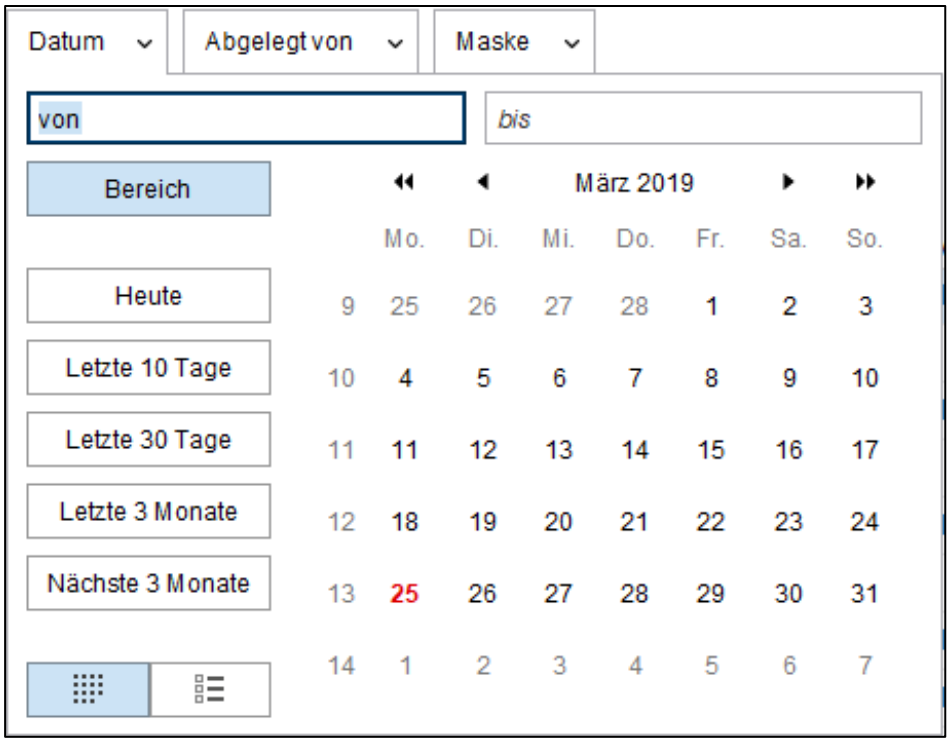


Abb. 8: Bereichseingabe mit zwei Eingabefeldern

Bei einer nicht als Datum erkennbaren Eingabe bleibt der Filter nach der Betätigung der ENTER-Taste geöffnet und das Eingabefeld wird rot eingefärbt, um den Fehler anzuzeigen.

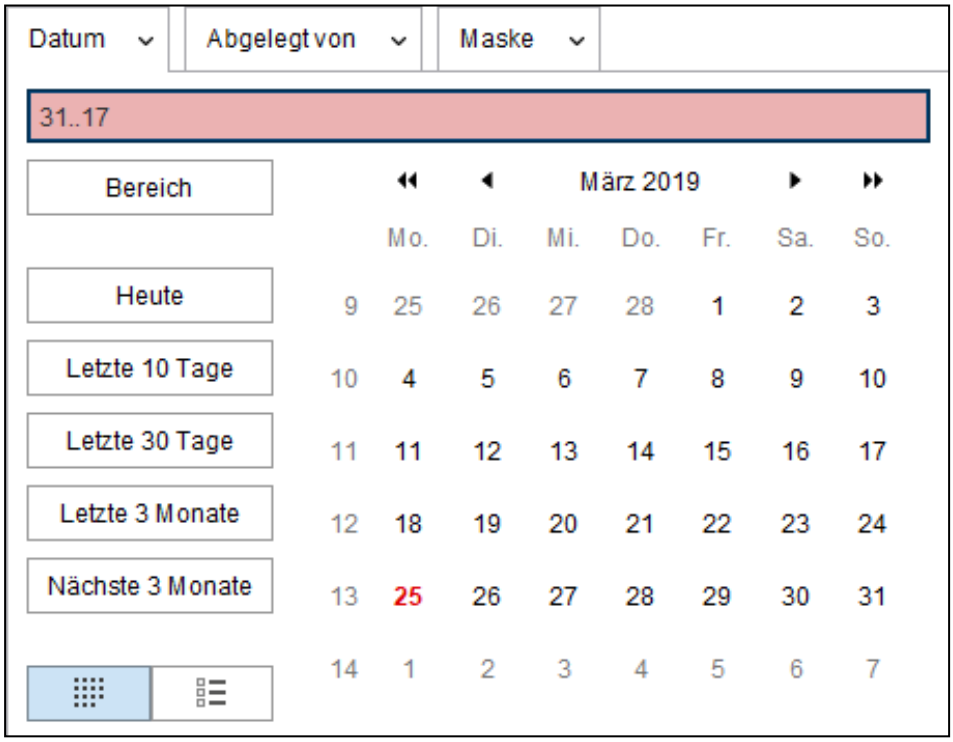


Abb. 9: Fehlerhafte Datumseingabe

Es können auch relative Zeiträume eingegeben werden. Dazu verwendet man Minus und Plus als Vorzeichen für einen Zeitraum in Tagen, ausgehend vom aktuellen Datum. Die Eingabe„-2“ sucht alles von vorgestern um 0 Uhr bis heute Ende des Tages. Mit einem angehängten „m“ können Monate verwendet werden. „+1m“ sucht alles von heute 0 Uhr bis einen Monat in der Zukunft (28, 29, 30 oder 31 Tage, je nach Monat und Jahr).



Abb. 10: Relative Suche

Links befinden sich fünf Schaltflächen, die häufige Zeiträume direkt anbieten. Der Zeitraum wird nach der Auswahl auch in das Eingabefeld eingetragen, um die verwendete Syntax anzuzeigen.

Rechts wird ein Kalender zur Datumsauswahl angezeigt. Das aktuelle Datum ist rot hervorgehoben. Über die Pfeile über dem Kalender kann in Monaten und Jahren vor- und zurückgesprungen werden. Man wählt per Mausklick einen einzelnen Tag. Mit SHIFT-Mausklick auf einen zweiten Tag ist der dazwischenliegende Zeitraum ausgewählt.

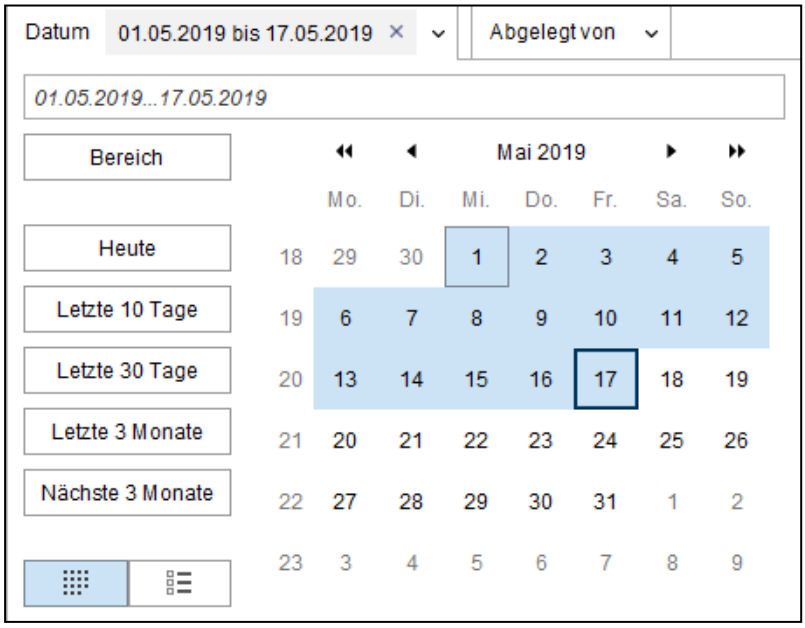


Abb. 11: Ausgewählter Zeitraum im Kalender

Links unten befinden sich zwei weitere Schaltflächen, um zwischen dem Kalender und der Jahresübersicht in Listenform umzuschalten. Diese zeigt alle vorhandenen Jahreswerte und jeweils rechts in Klammern die Anzahl der dazugehörigen Einträge. Sie können ein einzelnes Jahr (durch Mausklick in die Liste) oder mehrere Jahre (per Mausklick auf die Kontrollkästchen) auswählen.

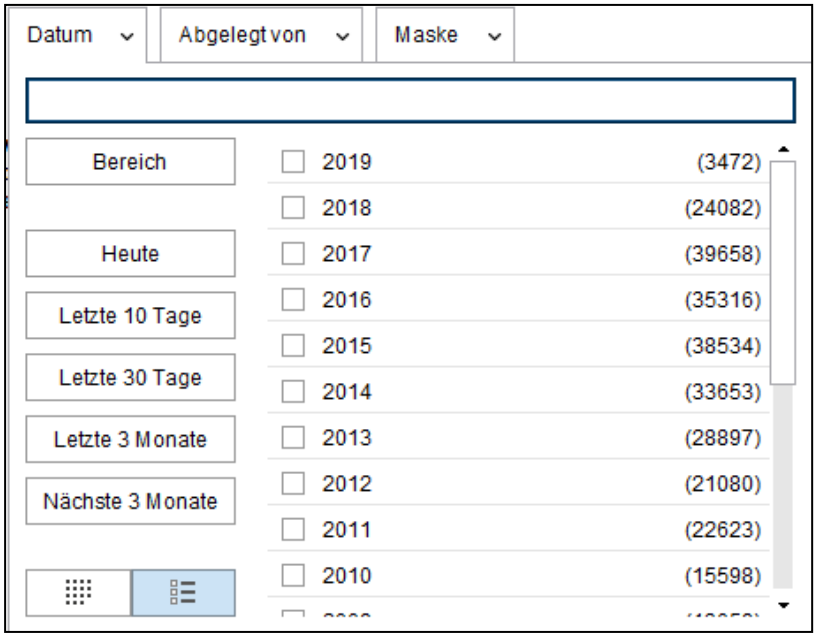


Abb. 12: Jahresübersicht in Listenform

3.4.3 Abgelegt von/Bearbeiter

Die Filter für „Abgelegt von“ und „Bearbeiter“ gleichen denen von Maske und Eintragstyp. Als Besonderheit wird hier ganz oben der aktuell in ELO angemeldete Benutzer angezeigt. Damit ist es sehr einfach, Dokumente zu finden, die bestimmte Benutzer abgelegt haben, oder auch die eigenen Dokumente.

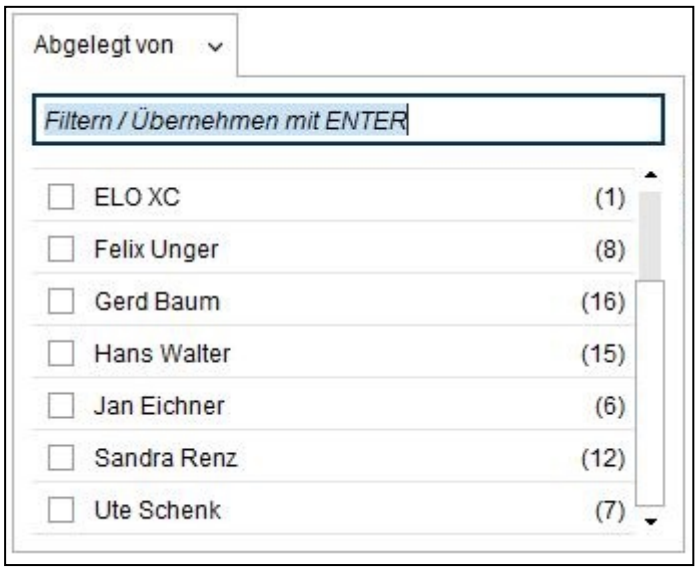


Abb. 13: Filter 'Abgelegt von'

3.4.4 Maske

Der Filter zeigt alle verwendeten Verschlagwortungsmasken und jeweils rechts in Klammern die Anzahl der dazugehörigen Einträge. Die Masken sind in alphabetischer Reihenfolge sortiert. Sie können eine einzelne Maske oder mehrere Masken über die Kontrollkästchen auswählen.

Durch Eingabe in das Filterfeld können Sie Search-as-you-type verwenden. Dabei wird Ihre aktuelle Eingabe sofort mit den möglichen Werten verglichen. Die Liste wird dadurch so eingegrenzt, dass nur noch die Masken angezeigt werden, die den eingegebenen Text enthalten. Es spielt keine Rolle, an welcher Stelle der Text steht. Im folgenden Beispiel wird daher durch Eingabe von „Re“ sowohl „Rechnung“ als auch „SupportRegister“ gefunden. Die Trefferstelle wird hellgelb hinterlegt.

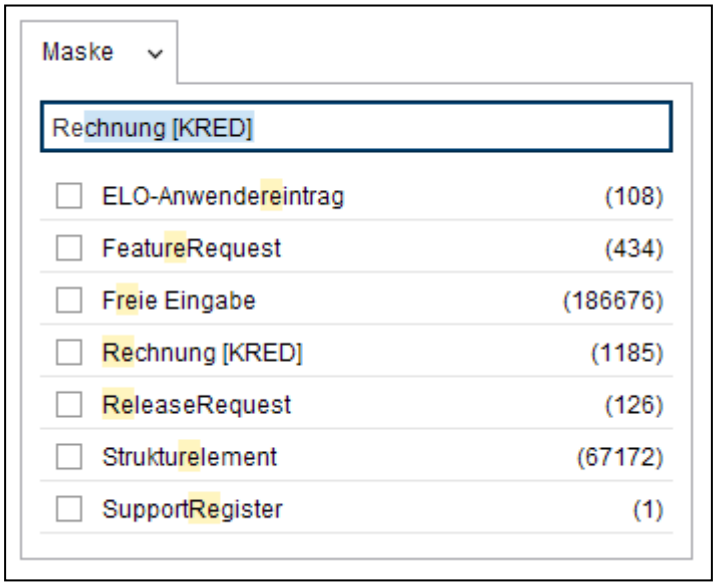


Abb. 14: Beispiel für Search-as-you-type

Bei der Eingabe wird außerdem eine Autovervollständigung vorgeschlagen. Dazu wird die Eingabe als Anfang des Maskennamens verstanden und die passende Maske mit den meisten Einträgen als weiterer Text in das Eingabefeld eingetragen und markiert. Mit ENTER können Sie die Autovervollständigung akzeptieren und übernehmen. Oder Sie tippen weiter, wenn Sie eine andere Maske suchen.

3.4.5 Alle Indexfelder

Dieser Filter durchsucht alle Indexfelder in allen Verschlagwortungsmasken. Er verwendet dabei die gleichen linguistischen Regeln wie das Eingabefeld der iSearch und stellt einen einfachen Weg dar, um zwischen der Suche im Volltext (über das Eingabefeld) und in der Verschlagwortung (mit diesem Filter) gezielt zu unterscheiden.

Wenn Sie zusammenhängende Begriffe, numerische Werte oder Datumsangaben finden möchten, ist die Suche in einem spezifischen Indexfeld besser geeignet.

3.4.6 Indexfeld auswählen

Filter für bestimmte Indexfelder können manuell hinzugefügt werden. Dazu wählt man im Dialog zuerst die jeweilige Maske und dann die Indexfelder aus. Wenn Sie die Suche bereits mit einem Maskenfilter auf eine Maske begrenzt haben, ist diese im Dialog vorausgewählt.

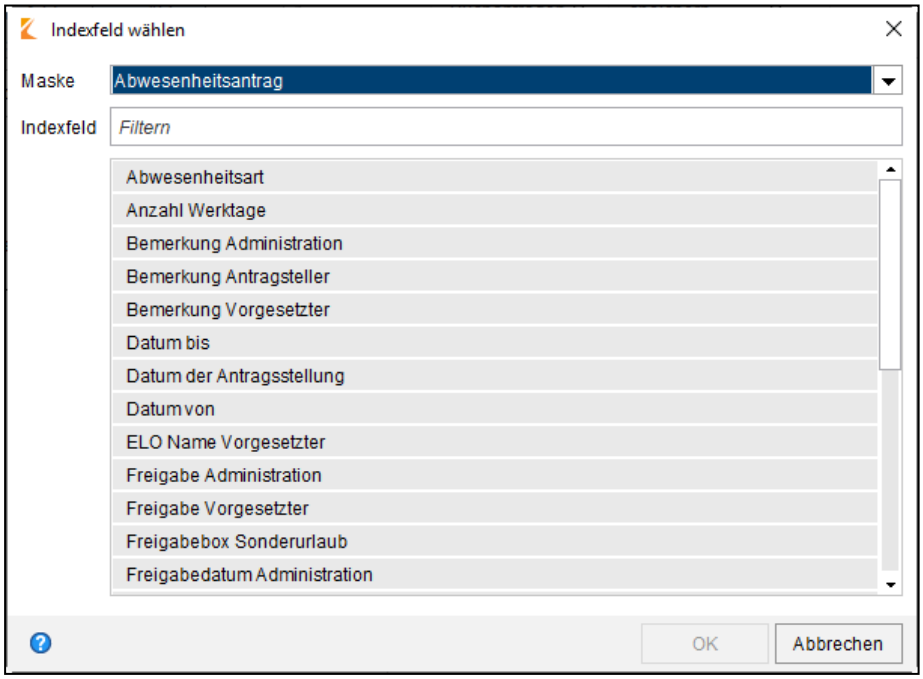


Abb. 15: Filter für Indexfeld einblenden

Information: Die Auswahl der Felder erfolgt für eine einfache Benutzung über die Maske. Der Indexfeld-Filter an sich ist aber maskenübergreifend, wenn die Indexfeldvorlage des ausgewählten Indexfelds auch in anderen Masken verwendet wird. Um die Suche auf eine Maske einzugrenzen, ist ein Maskenfilter notwendig.

Der Filter für ein Indexfeld kann mit Suchregel- und Suchmuster-Methoden oder kompletten Werten arbeiten. Dies ist eine Konfiguration der Indexfeldvorlage.

Die Option „Tokenisierung deaktivieren" (siehe auch Kapitel *4.5.8 Tokenisierung*) einer Indexfeldvorlage schaltet die Suchregeln und -muster für dieses Feld aus. Es erscheinen in der Liste des Filters die exakten Werte der Verschlagwortung. Die Suche arbeitet für dieses Feld dann ebenfalls in einem exakten Modus. Teilbegriffe werden in diesem Fall nicht berücksichtigt.

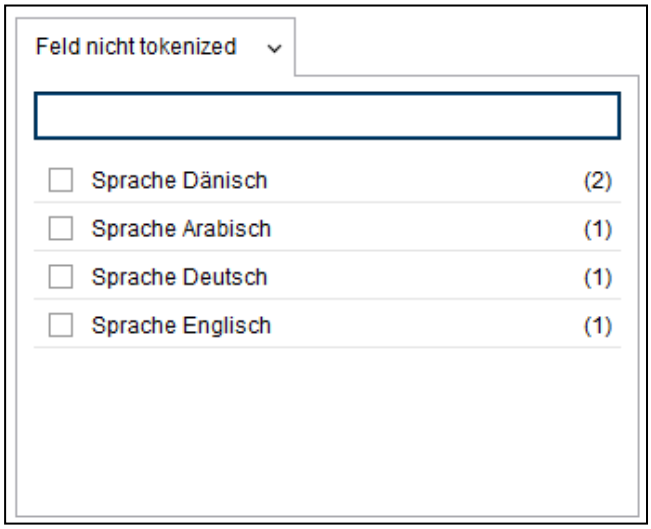


Abb. 16: Indexfeld nicht in Token aufgeteilt

Bei aktiven Suchregeln und -mustern erscheinen in der Liste des Filters nur einzelne Wörter, sowohl die Originale als auch die erzeugten Stammformen. Die Suchmöglichkeiten in diesem Filter entsprechen dann denen des Eingabefeldes der Suche.

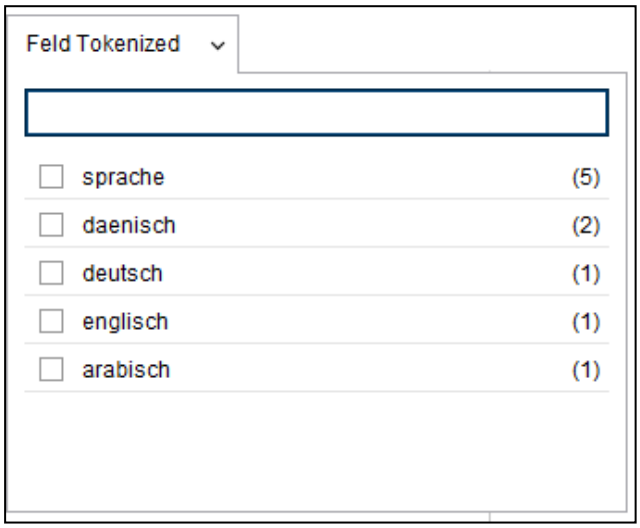


Abb. 17: Indexfeld in Token aufgeteilt

Einzelne Filter kann man auch über deren Kontextmenü ausblenden.

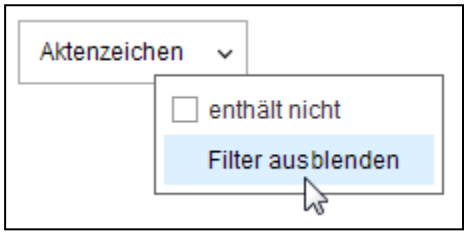


Abb. 18: Filter ausblenden per Kontextmenü

3.4.7 Numerische Indexfelder

Bei Indexfeldern, die als numerische Daten konfiguriert sind, zeigt die Liste des Filters Wertebereiche als Suchvorschläge an.

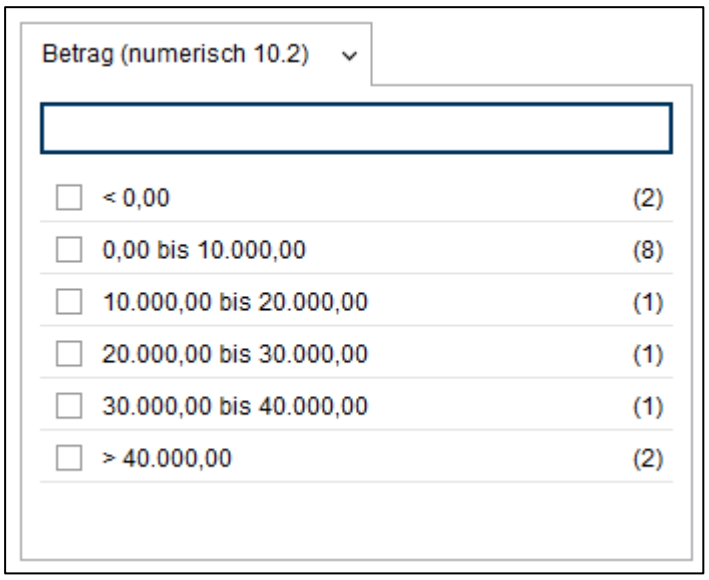


Abb. 19: Filter für numerische Indexfelder

Solche Bereiche können auch direkt mit drei Punkten als Bereichstrennzeichen in den Filter eingegeben werden, z.B. 100…900 für eine Suche von 100 bis 900. Außerdem ist eine Suche mit kleiner und größer als … möglich, z. B. >2500. Das Eingabefeld des Filters färbt sich bei der Eingabe (mit der ENTER-Taste) rot, wenn eine Filtereingabe nicht als numerischer Wert verwendbar ist.

3.4.8 Feed

Der Filter für die Suche im Feed dient der Suche über alle Feed-Inhalte. Dies sind sowohl von ELO Benutzern manuell erstellte Beiträge und Kommentare als auch automatisch generierte Einträge, wie zum Beispiel zu einer neuen Dokumentversion. Im Feed sind damit auch die Versionsnummern und Versionskommentare enthalten, sowohl zu der aktuellen Arbeitsversion eines Dokuments als auch zu älteren Versionen.

In der Liste des Feed-Filters werden die am häufigsten verwendeten Hashtags als Suchvorschläge angezeigt. Diese werden wie alle anderen Eingaben auch in diesem Filter nach den üblichen linguistischen Regeln im Textinhalt des Feed gesucht.

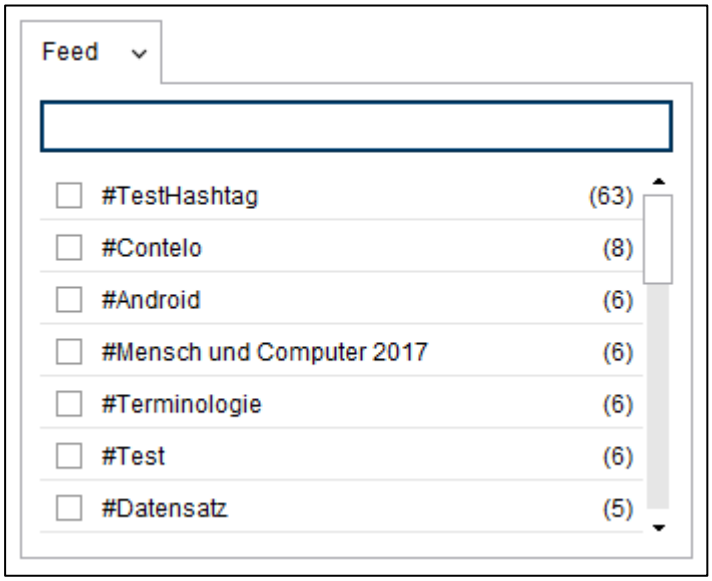


Abb. 20: Feed-Filter

3.4.9 Dateigröße

Der Filter für die Dateigröße eines Dokuments ist der Einzige ohne Eingabefeld. Die komfortable Auswahl des Größenbereichs erfolgt über eine Liste.

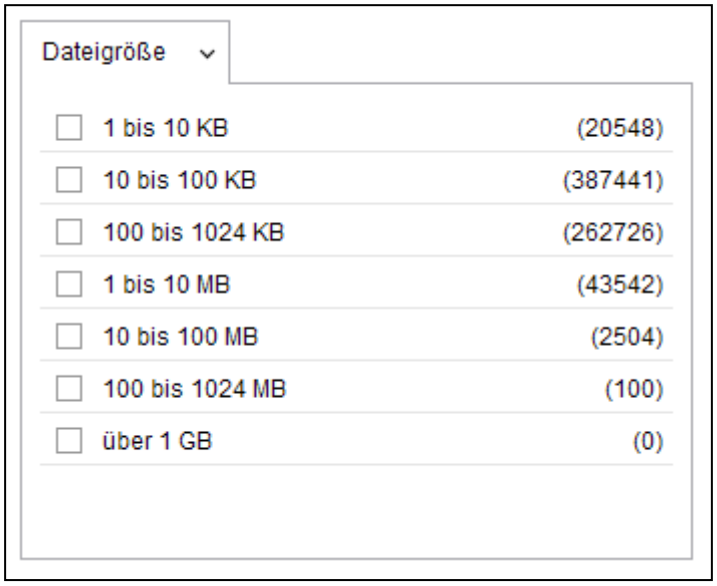


Abb. 21: Filter für Dateigröße

3.4.10 Suche im aktuellen Ordner

Die Suchoption „Nur aktueller Ordner" befindet sich in der Multifunktionsleiste und grenzt die ELO iSearch auf den zuletzt benutzten Ordner ein – im Tooltip wird kann der aktuelle Ordner angezeigt werden.

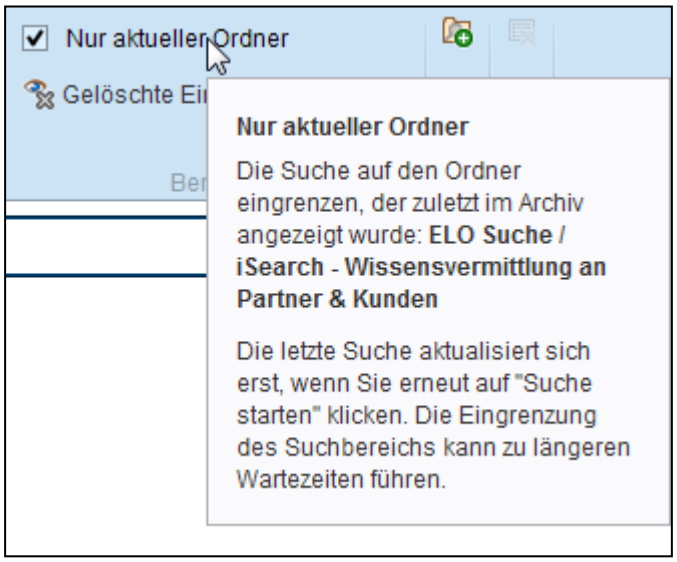


Abb. 22: Suche im aktuellen Ordner

Da es sich bei dieser Einstellung um eine nachträgliche Filterung der Treffer aus dem Suchindex handelt, verlangsamt dies die Suche deutlich. Es ist daher sinnvoll, diese Einstellung nur dann einzusetzen, wenn bereits eine relativ kleine Treffermenge über andere Suchkriterien sichergestellt ist. Die Suchfilter bieten sich dafür an.

3.5 Suche speichern

In der ELO iSearch ist es möglich, Sucheingaben zu speichern. Verwenden Sie dazu die Funktion „Suche speichern" (*Multifunktionsleiste > Recherchieren > Bereich Favoriten*).

Mit dem Speichern einer Suche halten Sie die gesamte ELO iSearch-Sucheingabe fest, inklusive der Auswahl der Filter und Filterwerte. Die Einstellung „Nur aktueller Ordner" (in der Multifunktionsleiste) wird dabei nicht abgespeichert, da diese Auswirkungen auf die Suchperformance hat und daher nur in Ausnahmefällen verwendet werden sollte.

Sie können sich zum Beispiel einen Suchfavoriten für die Recherche nach aktuellen E-Mails anlegen. Dann müssen Sie nur noch das jeweilige Suchwort eingeben und haben Maske und Zeitraum bereits richtig vorbelegt.



Abb. 23: Suchfavorit für aktuelle E-Mails

Beim Speichern der Suche werden zwei Optionen angeboten: Das Anlegen eines neuen Suchfavoriten oder die Anpassung der Standardfilter.

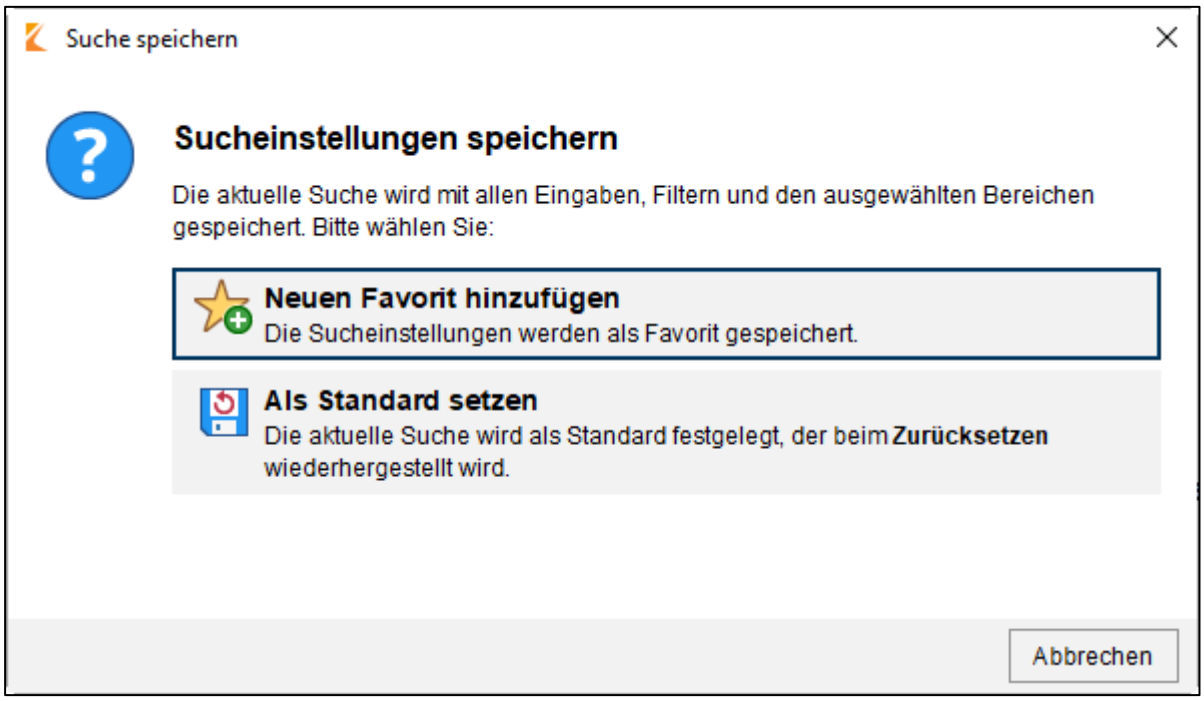


Abb. 24: Dialog 'Suche speichern', Auswahl 'Neuen Favorit hinzufügen'

3.5.1 Standardfilter

Die in Kapitel *3.4 Filter* beschriebenen Standardfilter können man per „Suche speichern" an die jeweiligen Vorlieben angepasst werden. Der somit für Sie definierte persönliche Standard wird beim Start des Clients verwendet und lässt sich jederzeit über die Schaltfläche „Auf Standard zurücksetzen" wiederherstellen.

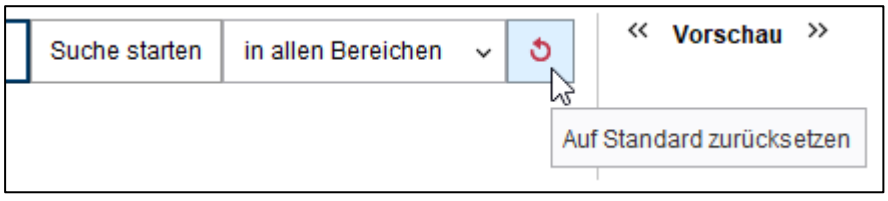


Abb. 25: Schaltfläche 'Auf Standard zurücksetzen'

Wird gerade ein Suchfavorit verwendet und angepasst, wird zunächst auf den gespeicherten Zustand des Suchfilters zurückgestellt. In diesem Fall gelangt man mit einem zweiten Aufruf von „Auf Standard zurücksetzen" zurück zu den Standardfiltern.

3.5.2 Suchfavoriten

Suchfavoriten geben Ihnen die Möglichkeit, verschiedene wiederkehrende Rechercheaufgaben zu vereinfachen. Die Favoriten lassen sich in der Multifunktionsleiste einblenden. Diese erscheinen nach einem Neustart dann auch als Kacheln im Bereich „Mein ELO“, clientübergreifend im ELO Java Client und ELO Web Client. Beim Anlegen eines Suchfavoriten werden Sie als Benutzer aufgefordert, einen Namen bzw. eine eindeutige Bezeichnung zu vergeben.

Haben sie gerade einen Suchfavoriten in Verwendung und speichern dann die Suche, so bietet Ihnen der Dialog zusätzlich die Möglichkeit, auch den aktuellen Suchfavoriten anzupassen:

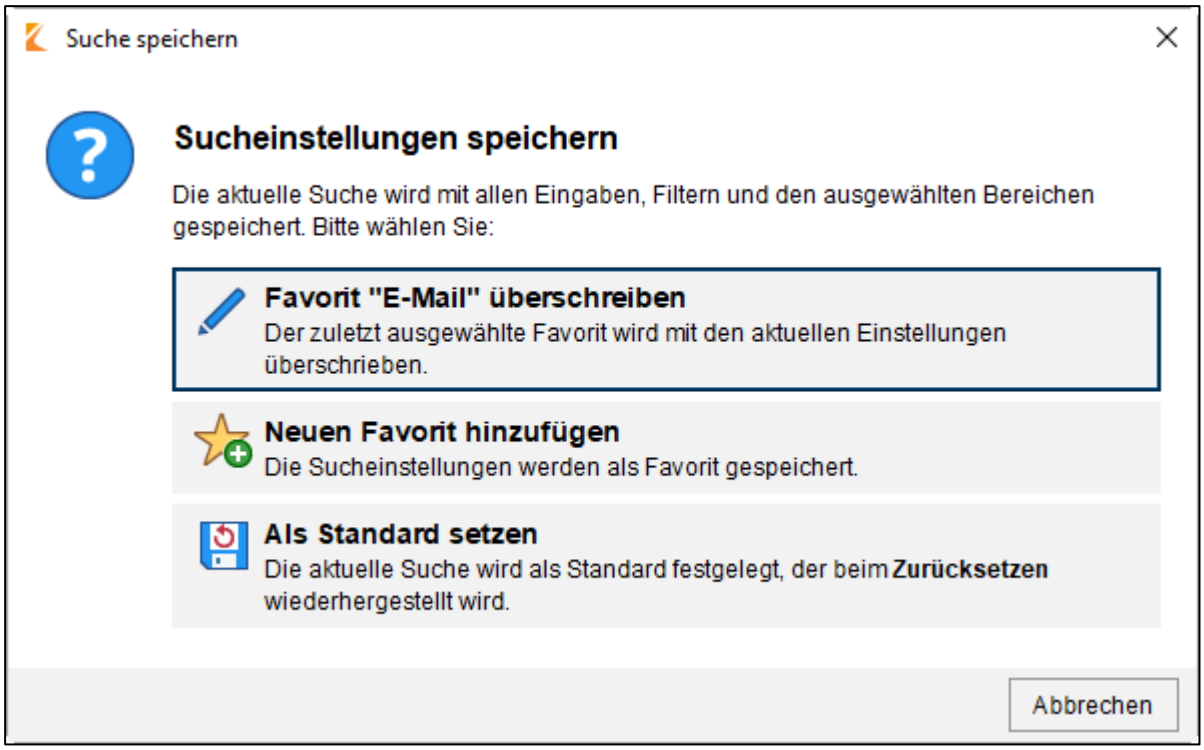


Abb. 26: Dialog 'Suche speichern', Auswahl 'Favorit "E-Mail" überschreiben

3.5.3 Suchfilter administrieren

Als Administrator besteht die Möglichkeit, Suchfilter nicht nur für sich persönlich, sondern auch für andere, für Optionsgruppen oder für alle Benutzer global im Archiv zu definieren. Die Funktion „Favoriten verwalten“ enthält in diesem Fall am oberen Rand eine Leiste zur Auswahl, wessen Favoriten verwaltet werden sollen.

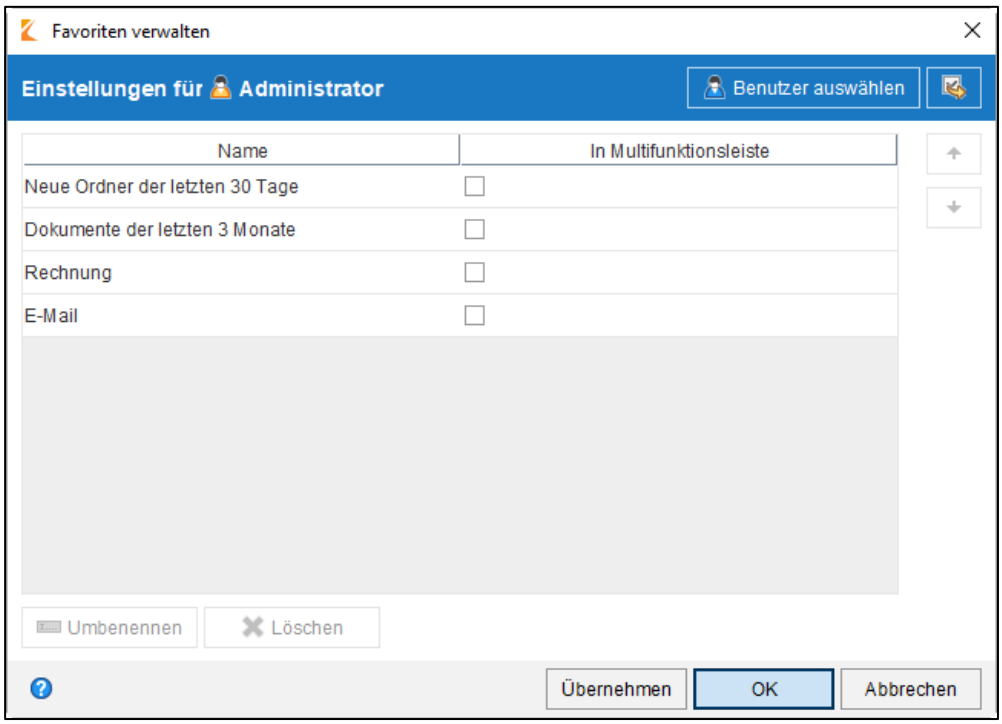


Abb. 27: Funktion 'Favoriten verwalten'

Man kann damit bei einem Benutzer oder bei einer Optionsgruppe Suchfavoriten umbenennen, löschen oder in die Multifunktionsleiste einblenden. Mit der Schaltfläche „Einstellungen auf andere Benutzer übertragen“ können im Dialog ausgewählte Suchfavoriten an andere Benutzer, Optionsgruppen und global verteilen. Hierfür wird auch die Einstellung der Standardfilter im Dialog eingeblendet, wenn diese beim ausgewählten Benutzer gespeichert wurde.

3.6 Trefferliste

Als Ergebnis der Suche wird eine Tabelle (Trefferliste) mit Einträgen angezeigt, die der Suchanfrage entsprechen. Die Sortierung der Einträge erfolgt chronologisch absteigend nach ihrem Ablagedatum, d. h. die neuesten Dokumente stehen in der Trefferliste oben. Diese Sortierung der Trefferliste kann jederzeit angepasst werden, entsprechend der jeweiligen Bedürfnisse und der für das Suchergebnis relevanten Werte und Informationen.

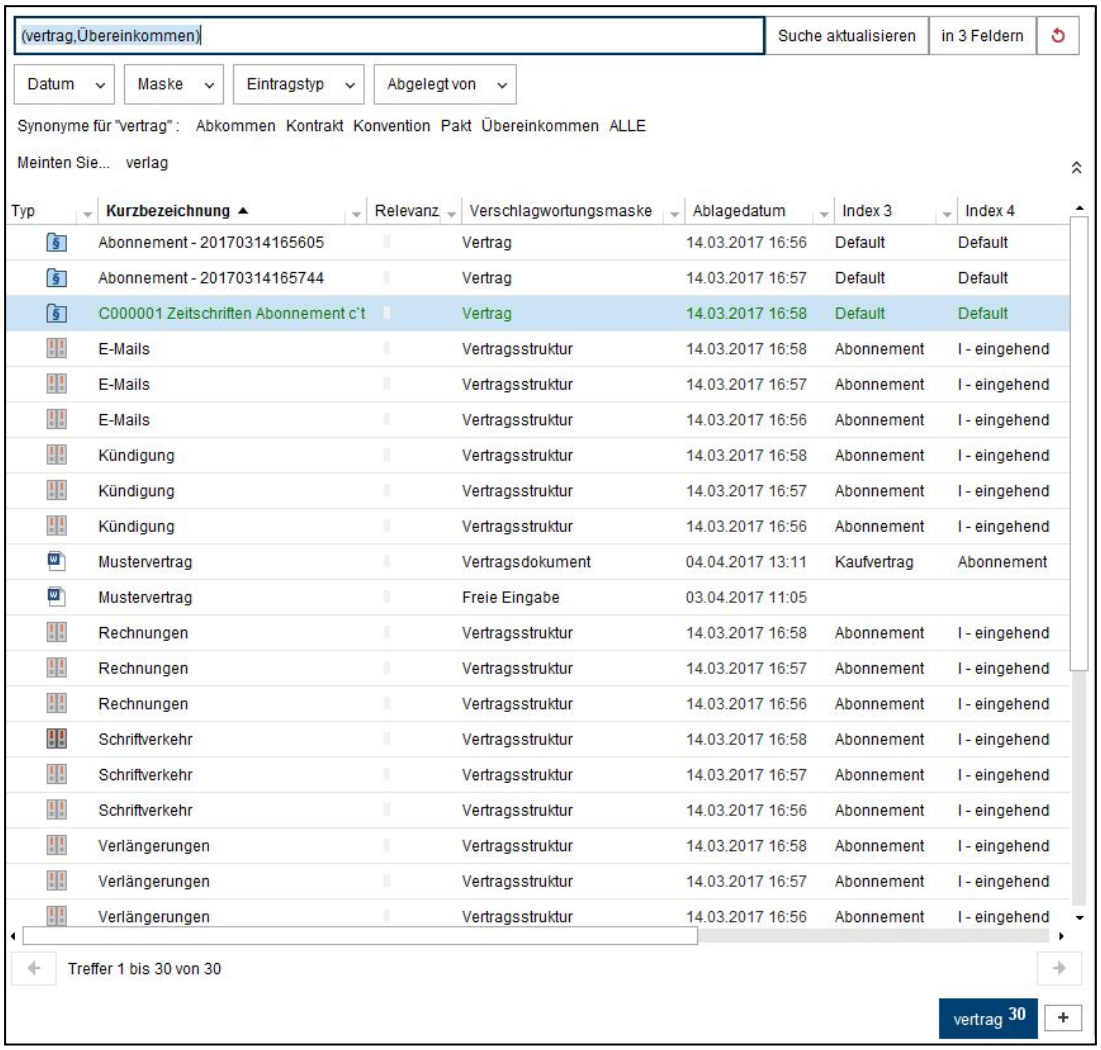


Abb. 28: Beispiel einer Tabellenansicht in der Trefferliste

Information: Bei der Anzeige der Fundstellen in der Tabellenansicht (siehe entsprechende Spalte), werden alle zur Sucheingabe passenden Fundstellen für das gefundene Dokument angezeigt, auch wenn eine konkrete Fundstelle im Suchbereich (siehe Abschnitt *3.3 Suchbereich/Durchsuchte Felder*) ausgeschlossen worden war. Das betreffende Dokument wurde jeweils nur über den eingestellten Suchbereich gefunden, jedoch kann es für den Benutzer hilfreich sein, dass er auf weitere im Dokument vorkommende Treffertexte aus allen Feldern hingewiesen wird.

3.6.1 Treffermenge und -seiten

Um die Anzeige der Treffermenge weiter zu beschleunigen, gibt es einen Parameter, der die maximale Anzahl an Treffern eingrenzt, die angezeigt werden sollen. Dieser Parameter ist serverseitig konfigurierbar und kann den Anforderungen entsprechend angepasst werden. Hierfür ist die entsprechende Berechtigung notwendig.

Wird diese Anzahl bzw. dieser Parameterwert überschritten, werden lediglich ältere Dokumente nicht im Suchergebnis angezeigt. Durch Verwendung eines Datumsfilters kann man gezielt auch auf diese Dokumente zugreifen oder aber diese ganz bewusst ausschließen.

Der ELO Java Client zeigt die Treffer aufgeteilt in einzelne Trefferseiten an. Dies ist notwendig, da die Aufbereitung der Ergebnisse linear von der Anzahl der Treffer abhängt. Mit den Pfeilen kann seitenweise weiter- oder zurückgeblättert werden.

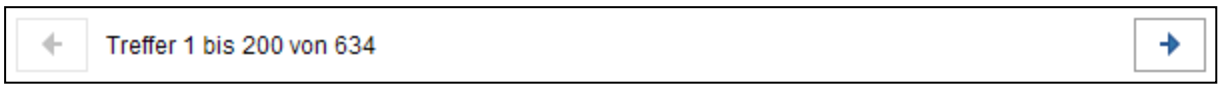


Abb. 29: Suchergebnis auf mehreren Trefferseiten

Beachten Sie: Die angezeigte Gesamttrefferanzahl ist nur eine Schätzung. Die Berechtigungen im Archiv können sich durch fachliche Anwendungen zur Laufzeit ändern (z. B. das Ändern von Berechtigungen oder das Ablegen von neuen Dokumenten im Archiv).

Der Indexserver filtert daher alle Einträge nach, wodurch die angezeigte Treffermenge geringer oder größer ausfallen kann. Der Client passt den angezeigten Wert an, sobald das Trefferende erreicht und die Nachfilterung abgeschlossen ist. Auch die Suche im aktuellen Ordner verwendet eine nachträgliche Filterung. Hier weichen die Werte deutlich ab, da sich die Schätzung immer auf das ganze Archiv bezieht.

3.6.2 Darstellung als Liste, Kachel oder Baum

Der ELO Java Client bietet bei der Darstellung der Suchtreffer neben der Tabellenansicht auch eine Listendarstellung, eine Kachel- oder eine Baumansicht an (*Multifunktionsleiste > Ansicht > Bereich Darstellung*).

Die Listendarstellung ist eine besonders kompakte Darstellung der Suchergebnisse, die sich gut für eine übersichtliche Anzahl von Treffern eignet und dadurch viel Platz für die Vorschau der Dokumente auf dem Bildschirm lässt.

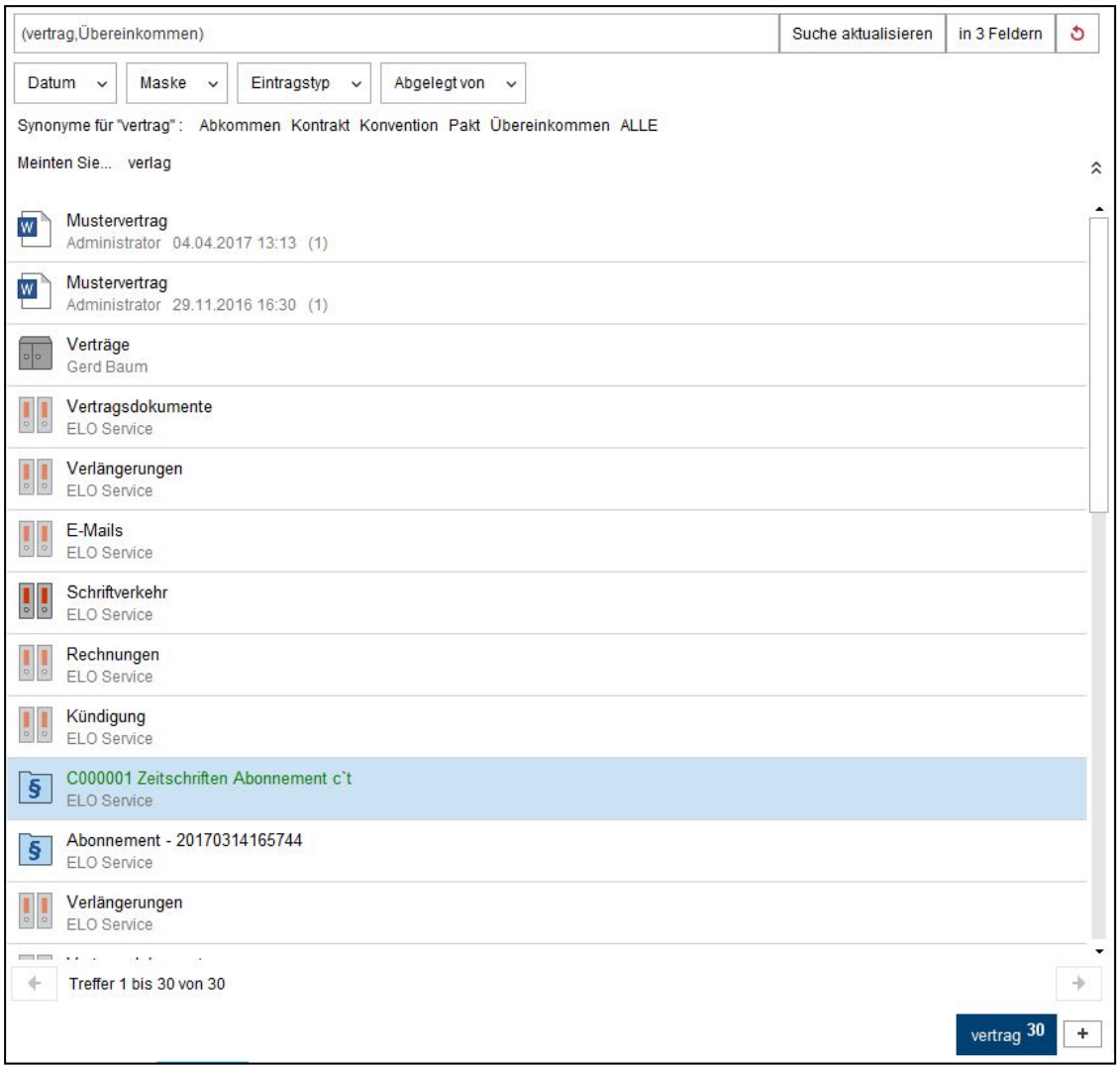


Abb. 30: Beispiel einer Listendarstellung

Die Kacheldarstellung eignet sich vor allem für die Anzeige von Miniaturansichten. Zu beachten ist hierbei, dass diese vom Client erzeugt werden, dafür also alle Dokumente zunächst vom Server geladen und bearbeitet werden müssen. Diese Darstellung eignet sich für überschaubare Treffermengen im lokalen Netzwerk.

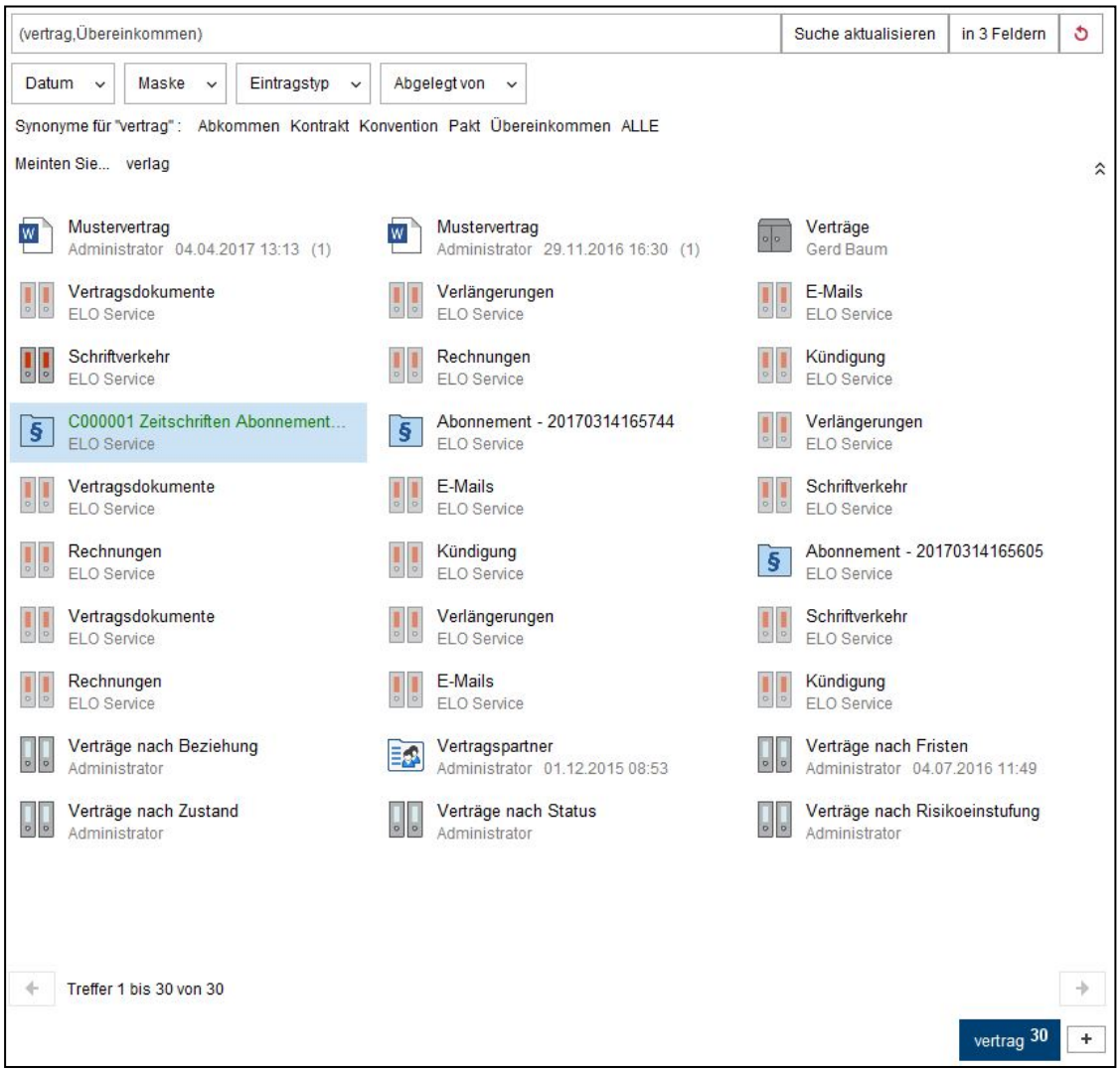


Abb. 31: Beispiel einer Kacheldarstellung

Die Baumdarstellung ermöglicht es, die Treffer nach auswählbaren Kriterien in einer Baumstruktur zu ordnen. Eine vordefinierte Möglichkeit ist die Darstellung entsprechend der Archivstruktur. Es können auch eigene Strukturen anhand der Indexfelder der Einträge definiert werden. Für eine Maske „Rechnung“ wäre zum Beispiel eine Struktur nach Jahr und Kunde möglich, wenn die benötigten Werte in der Maske als Indexfelder vorhanden sind.

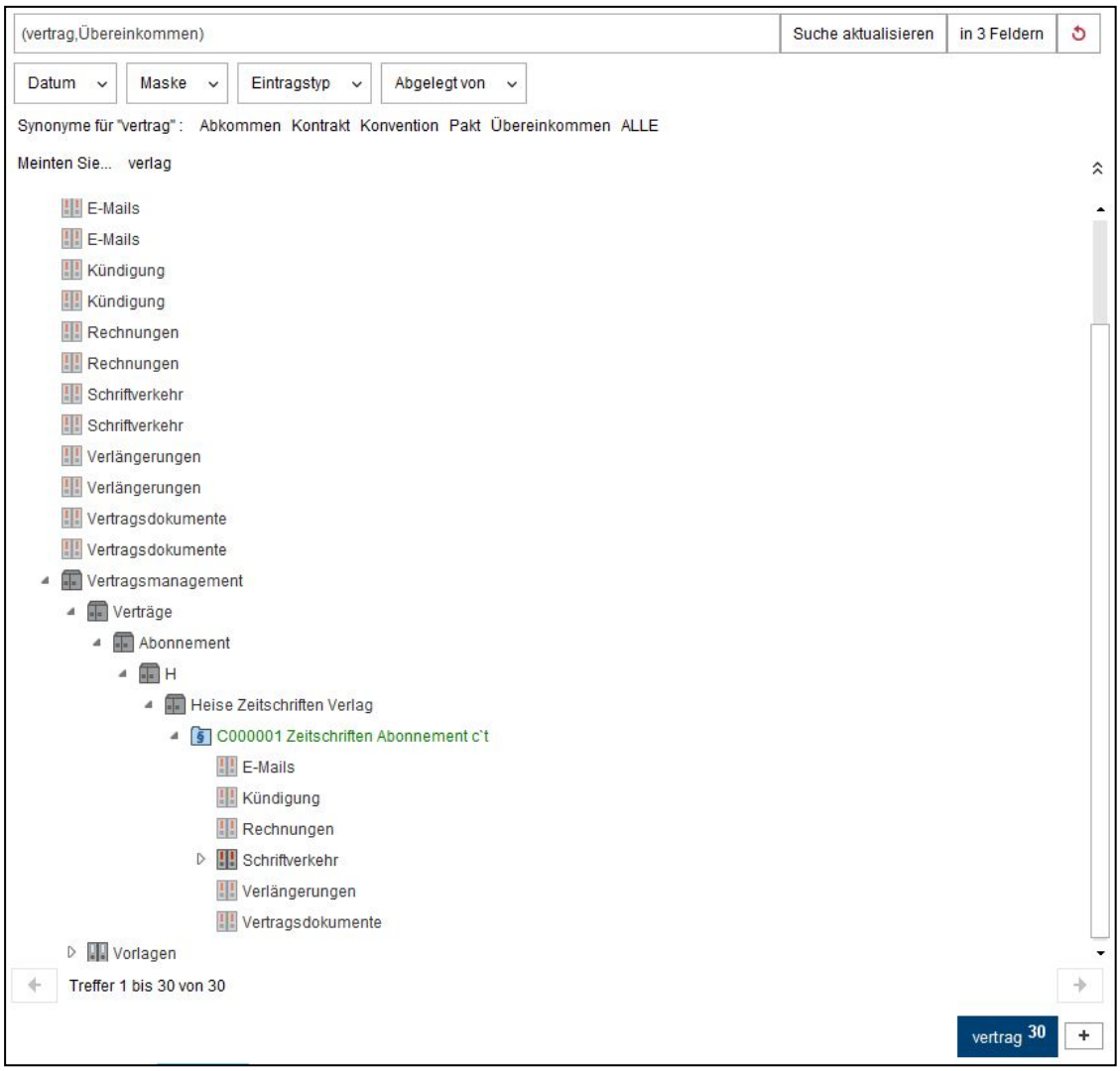


Abb. 32: Beispiel einer Baumdarstellung

Für eine Maske „Vertrag“, wie im oben dargestellten Beispiel, ist eine Struktur nach Anfangsbuchstabe Vertragspartner („H“) und Vertragsnummer sowie Vertragspartner („*C000001 Zeitschriften Abonnement c´t*“) unter Umständen sinnvoller für den Benutzer. Immer vorausgesetzt, die hierfür benötigten Werte sind in der Maske als Indexfelder vorhanden.

4 Technische Beschreibung

4.1 Architektur

4.1.1 Elasticsearch

Die Elasticsearch ist eine verteilte, hoch skalierbare Open-Source-Such-Engine auf Basis von Apache Lucene. Mit der Elasticsearch ist es möglich, große Datenmengen zu speichern und extrem schnell zu durchsuchen. Die Kommunikation erfolgt dabei über ein RESTful Webinterface.

Durch die große Verbreitung können im Internet viele Grundlagendokumente und Vorgehensweisen nachgeschlagen werden. In diesem Dokument werden deshalb lediglich die Details herausgearbeitet, die relevant für den Einsatz im ELO Systemumfeld sind. Für Informationen darüber hinaus kann folgender Einstiegspunkt für weitere Recherchen empfohlen werden:

https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html

4.1.2 Lucene

Die Elasticsearch nutzt Apache Lucene als „core library". Mit der Elasticsearch kann jeder Index in mehrere Stücke (sogenannte „Shards") aufgeteilt werden. Ein Shard entspricht einem Lucene-Index. Dieser besteht physikalisch aus einem Ordner mit den zugehörigen Index-Dateien.

Die Shards können auf mehrere Server („Nodes") aufgeteilt werden. Es können „Replicas" gebildet werden. Dies ergibt dann einen „Cluster".

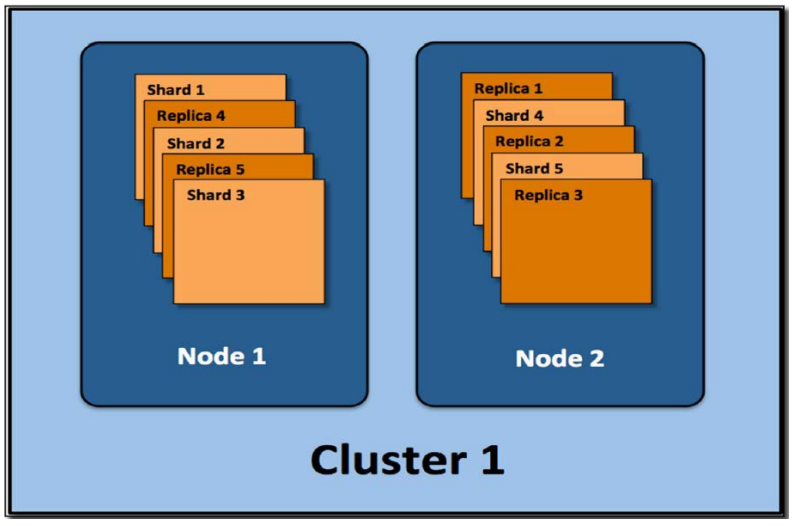


Abb. 33: Aufteilung durch die Elasticsearch in Cluster, Nodes, Replicas und Shards

4.1.3 Sprachunterstützung

Die Elasticsearch verwendet sowohl bei der Indexierung als auch bei der Suche eigene Sprach-Analyzer. Dadurch werden sprachspezifische Suchregeln und -muster optimiert. Bei den ELO Spezifika müssen Sie daher beachten, dass die Archivsprache (Auswahl bei der Installation) einen maßgeblichen Einfluss auf die Suche besitzt.

Eine Liste der unterstützten Sprachen im Kontext der Suchregel- und Suchmuster-Funktionen können Sie dem ELO Partner Portal oder der Statusseite entnehmen. Zum Zeitpunkt dieser Dokumenterstellung unterstützt die Elasticsearch in ELO folgende Sprachen:

> Deutsch, Englisch, Französisch, Spanisch, Italienisch, Portugiesisch, Arabisch, Dänisch, Schwedisch, Polnisch, Niederländisch, Tschechisch, Ungarisch, Rumänisch, Türkisch.

4.1.4 Verarbeitung mit ELO

Die Verarbeitung mit ELO kann dem folgenden vereinfachten Schema entnommen werden.

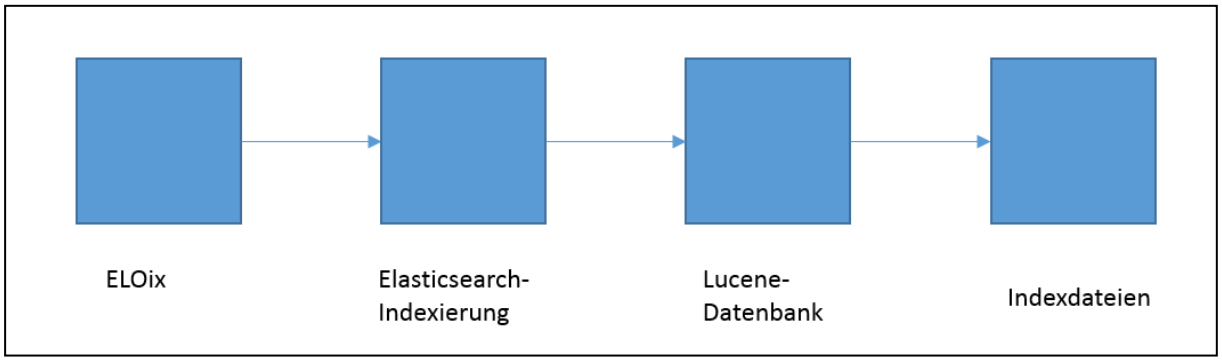


Abb. 34: Vereinfachtes Schema der Verarbeitung

Der ELO Indexserver übergibt die Daten an die Elasticsearch, die bei der Indexierung (und auch bei der Suche) eigene Sprach-Analyzer verwendet. Die Indexdaten werden in Lucene-Indexen gespeichert.

Für das Suchen im Index wird also der Zugriff auf:

- die Indexdateien,
- eine konsistente Lucene-Datenbank,
- ein funktionsfähiger Elasticsearch-Sprach-Analyzer,
- ein laufender ELO Indexserver (ELOix)

vorausgesetzt.

4.1.5 Sicherheit

Seit der ELO ECM Suite 10.2 erfolgt die Kommunikation zwischen dem ELO Indexserver (ELOix) und der Elasticsearch verschlüsselt (SSL). Dazu wird das „Search Guard Plugin“ verwendet.

4.2 Installation und Upgrade

4.2.1 ELO Server Setup

Die Installation oder ein Upgrade erfolgt über das ELO Server Setup. Dieses erzeugt – falls noch nicht vorhanden – die Konfigurationen, installiert das Elasticsearch-Programm und den Dienst, sowie ein Datenverzeichnis, in dem die Indexdaten abgelegt werden. Folgende Optionen können im ELO Server Setup gesetzt werden:

- Name des Dienstes
- Memory-Wert
- Port
- Datenverzeichnis
- Mit ELOenterprise auch weitere Elasticsearch-Server (siehe dazu auch Kapitel *5 Lizenzbestimmungen bei der Verteilung von Serverprozessen*)

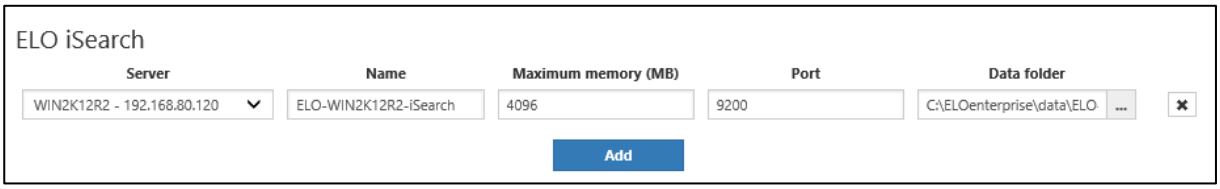


Abb. 35: ELO Server Setup, Bereich 'ELO iSearch'

4.2.2 Prüfungen nach der Installation

Nach der Installation können Sie zur Überprüfung als Tomcat-Administrator *http://<server>:<port>* (Standard-Port ist 9200) aufrufen. Bei Verwendung des Internet Explorers wird eine JSON-Datei ausgegeben, die in einem Editor geöffnet werden kann.

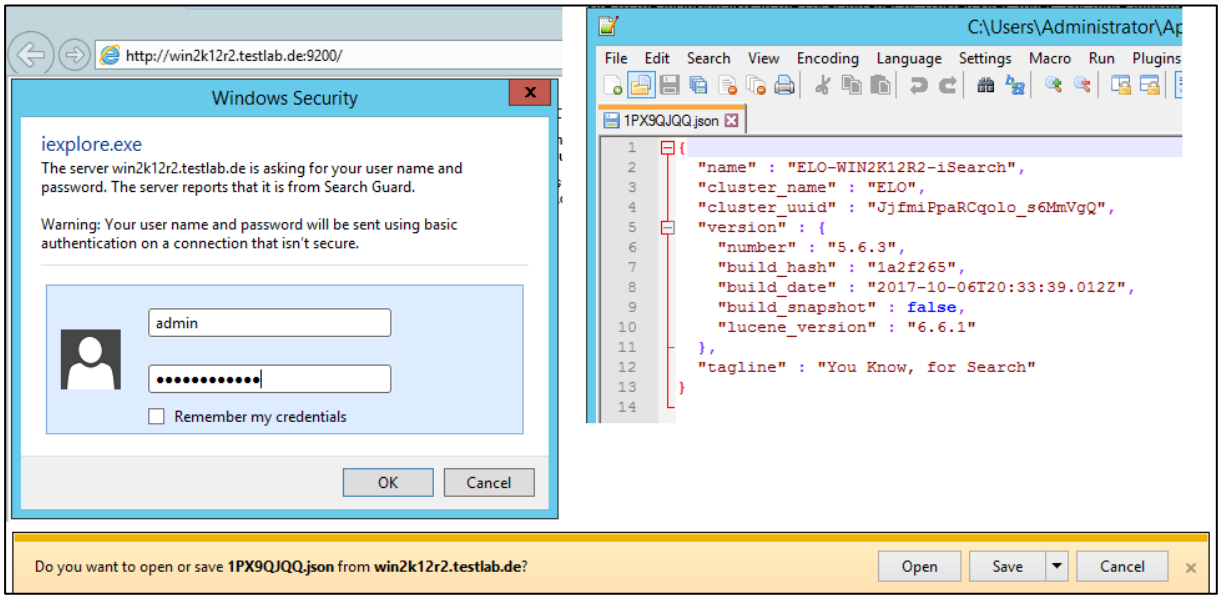


Abb. 36: Ausgabe der Elasticsearch-Basisdaten

Dadurch können Sie sich wichtige Basisdaten wie Version und Namen anzeigen lassen.

Über die ELO iSearch Configuration können Sie weitere Prüfungen vornehmen. Sie erreichen die Konfiguration über die Indexserver-Konfigurationsseite oder über folgenden Link (Anmeldung als Tomcat-Administrator):

http://<server>:<ixport>/ix-<ARCHIV>/manager/esconfig/#/iSearchConfig

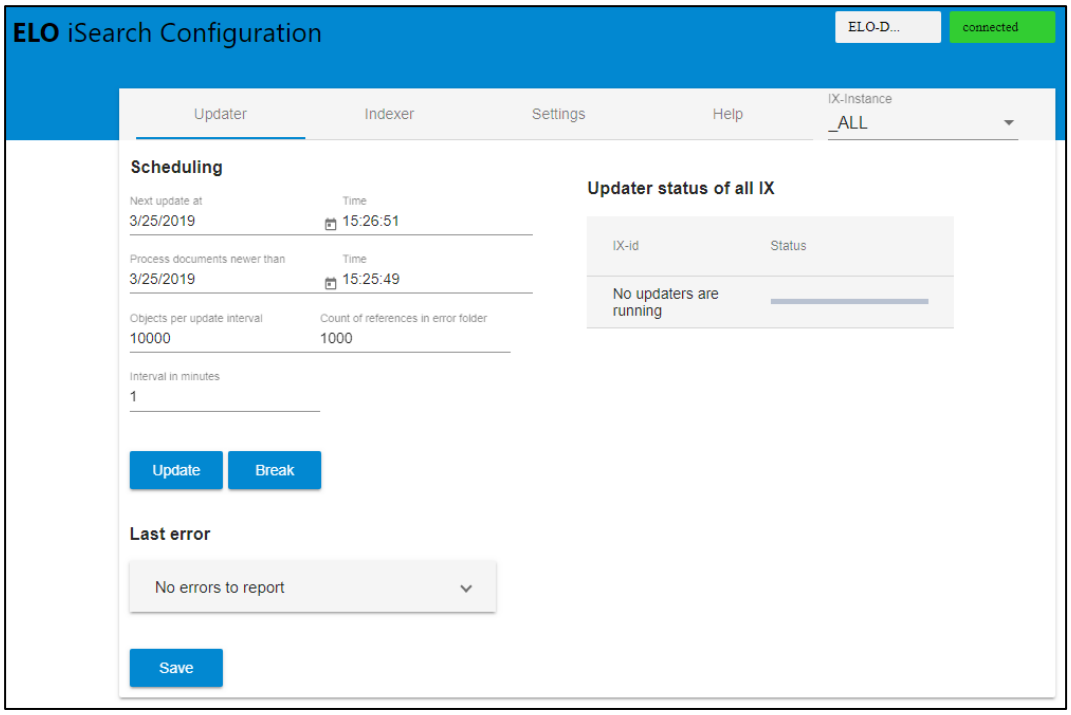


Abb. 37: Übersichtsseite ELO iSearch Configuration

4.2.3 Volltextindexaufbau nach Upgrade von älteren ELO Versionen

Unabhängig davon, ob ein Update von ELO 9 oder ELO 10 auf die Version 11 erfolgt, ist ein Neuaufbau (Re-Index) der Indexdaten notwendig, da sich die Indexdatenstruktur geändert hat.

Bei der Re-Indexierung wird zunächst der vorhandene Index gelöscht und dann neu aufgebaut. Alle Indexfelder werden sowohl „tokenized" (in Token aufgeteilt) als auch „not tokenized" (als Phrase) abgespeichert, so dass eine Änderung an dieser Indexfeld-Einstellung keinen Re-Index erfordert. Zusätzlich werden aus den durch den ELO Volltext-Prozess entstandenen „FT*.txt"-Dateien die Volltextinhalte ausgelesen und für die Indexierung verwendet.

Ein Re-Index wird über die ELO iSearch-Konfiguration im Tab „Indexer“ angestoßen.

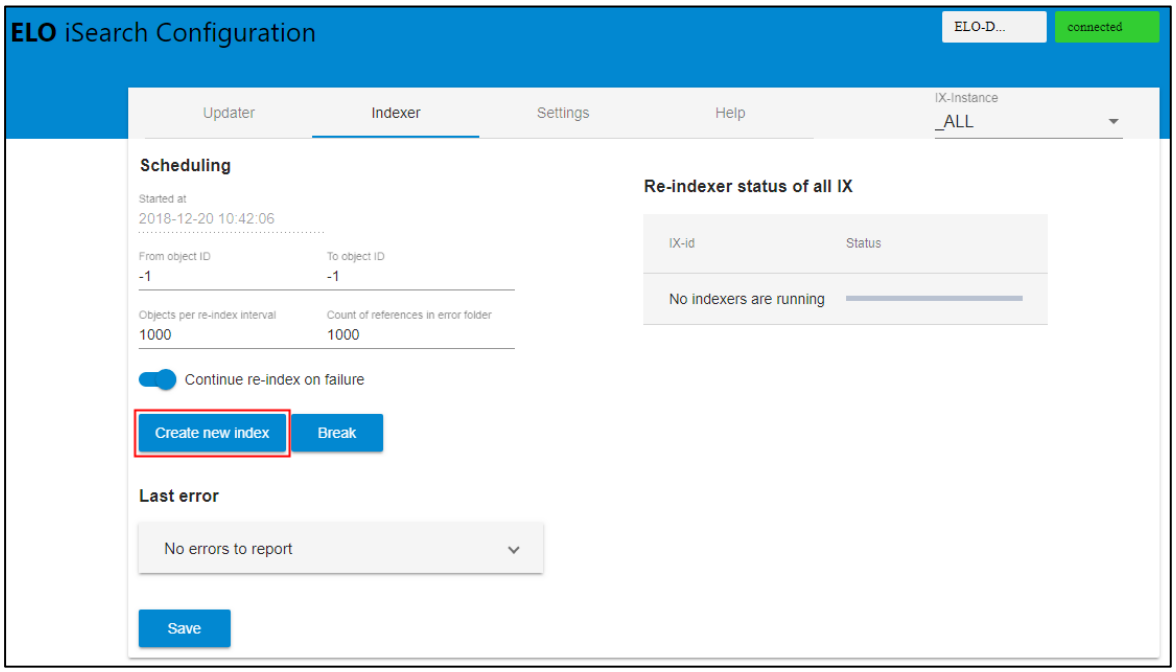


Abb. 38: ELO iSearch Configuration, 'Create new index'

Der Fortschritt wird nach einem Refresh der Seite angezeigt:

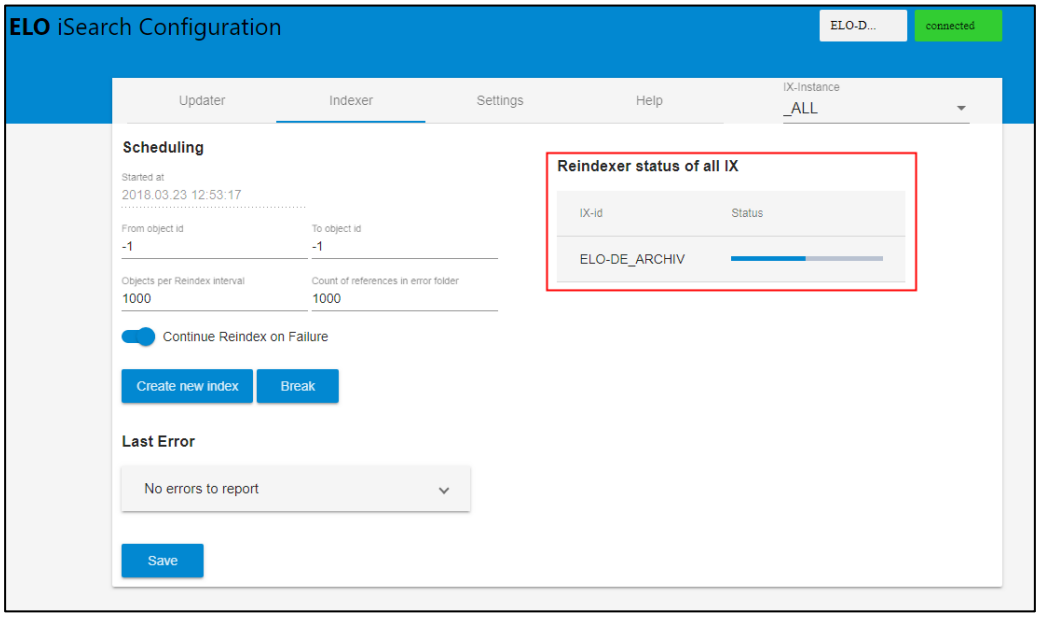


Abb. 39: ELO iSearch Configuration, Status der Re-Indexierung

4.2.4 Volltextindexaufbau im Hintergrund

Bei Upgrades ab der ELO Version 9.3 ist es möglich, über das ELO Server Setup mit der Funktion „Upgrade Index“ einen isolierten ELO Indexserver zu installieren, der den Elasticsearch-Index im Hintergrund aufbaut. Dies ist vor allem bei großen Systemen sinnvoll, bei denen die Re-Indexierung mehrere Tage dauert. Diese Funktion können Sie sowohl bei einem Update zu einer aktuelleren ELO Version verwenden als auch bei Änderungen in den Einstellungen der iSearch.

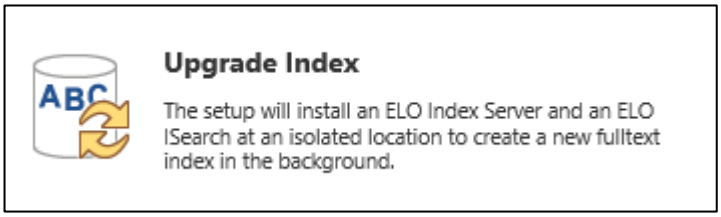


Abb. 40: ELO Server Setup, Upgrade Index

Diesen Index können Sie dann bei einem ELO Upgrade übernehmen, so dass direkt nach dem Upgrade die volle Suchfunktionalität zur Verfügung steht.

Details zum Verfahren können im Dokument „Volltextindexaufbau im Hintergrund“ nachgelesen werden, das im SupportWeb zur ELO iSearch zur Verfügung steht.

4.3 Sizing und Performance

Eine performante Suche stellt die Grundlage zur Akzeptanz des ELO Systems auf Benutzerebene dar. Daher müssen Sie neben der Optimierung der Datenbank-Performance auch regelmäßig (z. B. halbjährlich, in Abhängigkeit von der hinzukommenden Dokument- und Datenmenge) eine Optimierung der Elasticsearch durchführen. Die Notwendigkeit sollte zum einen auf Benutzerebene überprüft werden (durch Ausführung komplexerer Suchen), zum anderen aber auch auf Serverebene (durch gelegentliche Log-Datei-Analysen, siehe dazu auch Kapitel *4.4.1 Log-Datei*).

4.3.1 Allgemeine Aussagen zum Sizing

Die Performance der Elasticsearch wird durch folgende Faktoren beeinflusst:

- Arbeitsspeicher, sowohl für die JVM (Java Virtual Machine) der Elasticsearch, aber auch für das Betriebssystem allgemein
- Anzahl der Prozessoren/Kerne
- Speichertechnologie bzw. Speicher-Zugriffszeiten

Folgende generelle Aussagen zur Dimensionierung können gemacht werden:

- Arbeitsspeicher:
  - Für kleine Umgebungen sind die als Standard definierten 4 GB ausreichend.
  - Für mittlere Umgebungen (ca. 10 Mio. Dokumente) sollten ca. 20 GB eingeplant werden.
  - Für große Umgebungen gibt es eine Obergrenze von 31 GB je JVM, hier ist auf jeden Fall eine Verteilung auf mehrere Server (Cluster) sinnvoll.
- Prozessoren/Kerne:
  - Für jede Suche wird ein Thread benötigt, d. h. durch eine höhere Anzahl an Prozessoren/Kernen wird eine bessere Parallelisierung erreicht.
- Festplatten:
  - Schnelle Festplatten (SSD) machen sich signifikant bemerkbar. Besonders die Such-Performance profitiert von schnellen Zugriffszeiten auf den Speicher beim Laden der Indexdaten.

Auch bereits für mittlere Umgebungen empfehlen wir den Einsatz von ELOenterprise (siehe dazu auch Kapitel *5 Lizenzbestimmungen bei der Verteilung von Serverprozessen*) und die Implementierung der Elasticsearch auf einem dedizierten Server.

4.3.2 Optimierung der Indexierung

Den Bereich der Indexierung können Sie nur bedingt optimieren. Die Anzahl der Dokumente zur gleichzeitigen Indexierung wird vom ELO Indexserver (ELOix) bestimmt. Bei einem notwendigen Re-Index für große Dokumentenmengen raten wir dazu, einen eigenen Indexserver für diesen Task bereitzustellen (siehe dazu auch Kapitel *4.2.4 Volltextindexaufbau im Hintergrund*).

4.3.3 Optimierung der Suchgeschwindigkeit

Die Faktoren der Performance wurden schon in Kapitel *4.3.1 Allgemeine Aussagen zum Sizing* beschrieben. In großen Umgebungen ergeben sich weitere positive Einflüsse auf die Suchgeschwindigkeit durch eine Erhöhung der Anzahl der Shards je Index, wenn diese auf mehrere Nodes verteilt werden. Sie müssen dabei beachten, dass eine Anpassung der Shard-Anzahl einen Volltextindexaufbau (Re-Index) erfordert. Zur Konfiguration der Anzahl der Shards lesen Sie das Kapitel *4.5.5 Festlegung der Shard-Anzahl pro Index*.

4.3.4 Beispielhafte Systemsizings

In der folgenden Tabelle haben wir beispielhaft das Sizing diverser Kundenumgebungen zusammengestellt. In all diesen Kundenumgebungen ist die Elasticsearch auf dedizierten Servern implementiert.

Anzahl Dokumente	Anzahl Prozessoren/ Kerne	Arbeitsspeicher Elasticsearch JVM in GB	SSD	Bemerkungen zur Umgebung
8 Millionen	16	8	Ja	ELO 11 und ELO 12
65 Millionen, ca. 5 Millionen mit Volltext	4	8	Nein	ELO 10.1, geringer Workload
31 Millionen	6	26	Nein	ELO 10.2, hoher Workload
10 Millionen	2 x 4	2 x 8	Nein	ELO 10.2, moderater Workload, Cluster-Umgebung

4.3.5 Laufzeiten beim Volltextindexaufbau

Die Erfahrungswerte für den Volltextindexaufbau liegen im Bereich zwischen 20 Stunden für 1 Million Dokumente auf einer schwach ausgestatteten Umgebung ohne Tuning-Maßnahmen und 5 Stunden für 1 Million Dokumente auf einer leistungsfähigen Hardware mit Volltextindexaufbau im Hintergrund.

4.4 Betrieb

Für den Betrieb der Elasticsearch gibt es zwei wichtige Teilbereiche:

- Log-Dateien und deren Inhalt
- Erstellen von Auswertungen

4.4.1 Log-Datei

Standardmäßig liegen die Elasticsearch-Log-Dateien im ELO Installationsverzeichnis unterhalb des „logs"-Verzeichnisses.

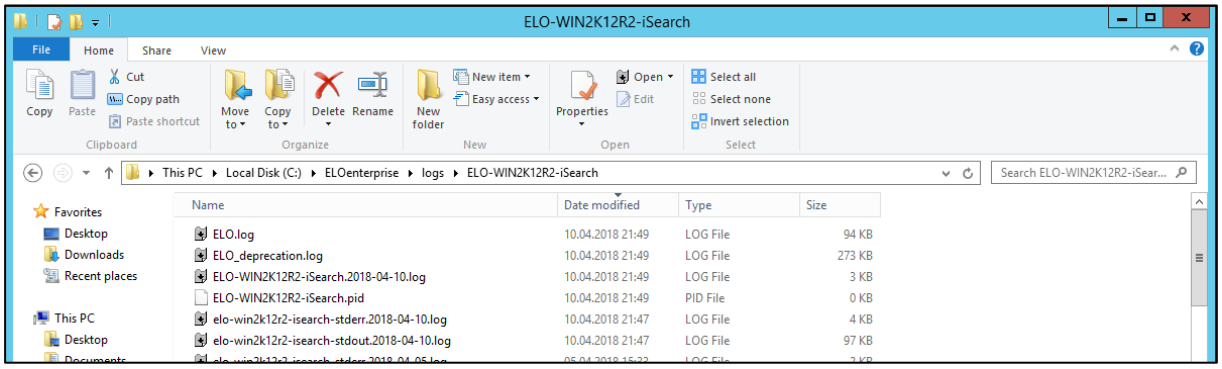


Abb. 41: Verzeichnis der Log-Dateien

Relevant ist hier die Log-Datei „ELO.log", bzw. „ELO-<datum>.log" für ältere Log-Dateien.

Im Einzelfall kann auch ein Blick in die Log-Datei des Indexservers („ix-<Archiv>.log") sinnvoll sein. Für ELO 11 und ELO 12 können Sie z. B. in der Log-Datei nach dem Begriff „queryTerm" suchen. Beispielhaft im Folgenden der Output für eine Volltextsuche nach dem Begriff „Test":

```
22:51:34,723 eloix-find-40 eloix-find-40 INFO  (ElasticClient.java:183) -
find(searchId=[(09A4500F-6BC6-4ECF-EFD6-A9E6B78C5A91)], queryTerm=Test,
sort=IDATE_DESC, highlightedText=false, resultField=true, relevance=true,
currentFolderId=0, searchIn=2)
```

4.4.2 Auswertungen

Diverse Status der Elasticsearch können Sie per Browser abfragen. Diese können Sie dann gegebenenfalls auch für Monitoring-Lösungen (wie z. B. Nagios) verwenden. Zur Anmeldung ist jeweils der Tomcat-Administrator notwendig. Ausgegeben wird jeweils eine JSON-Datei, die – wie z. B. in folgendem Screenshot – mit dem Firefox-Browser auch direkt angezeigt werden kann.

Die URL für „Health" lautet: *http://<server>:9200/_cluster/health?pretty=true*.

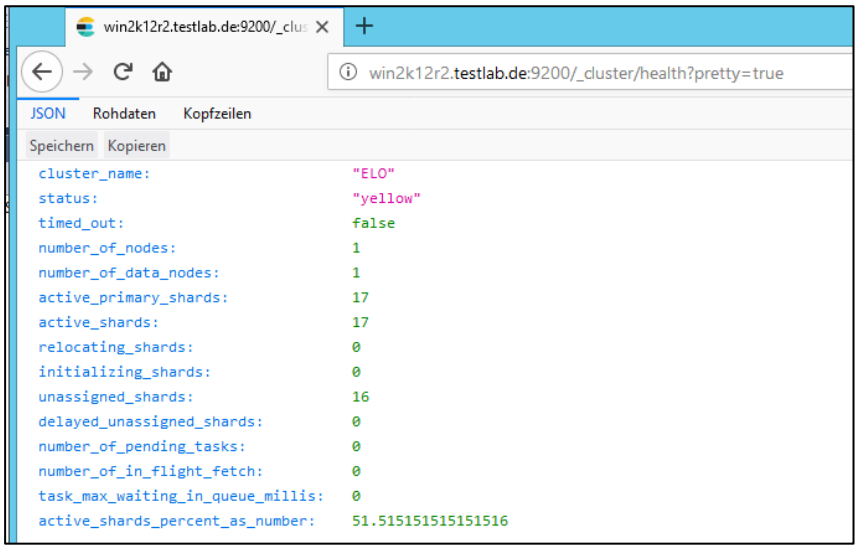


Abb. 42: Health-Zustand der Elasticsearch

Der Status „yellow" bedeutet, dass der Index funktionsfähig ist, die Replica-Shards jedoch nicht alle verfügbar sind. In einer Einzelserverumgebung, bei der die Standardeinstellung für Replicas (1) genutzt wird, ist der Status „yellow" normal.

Die URL für Indexverzeichnisnamen lautet: *http://<server>:<port>/_cat/indices*.

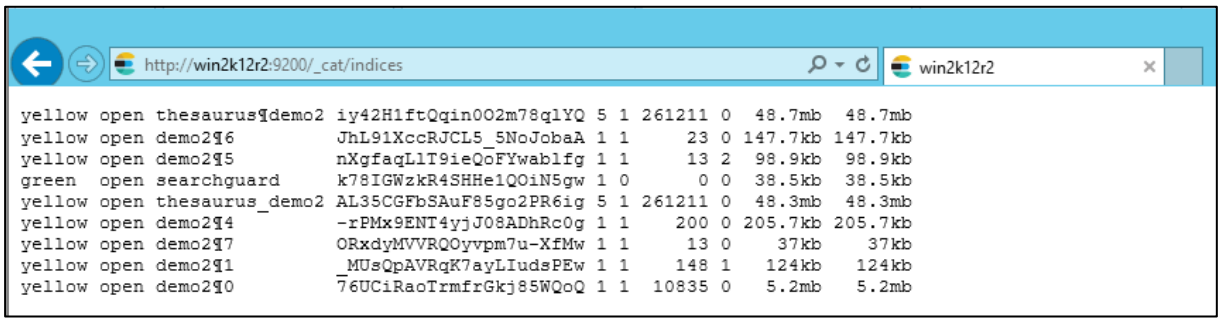


Abb. 43: Indexverzeichnis-Namen

Für jede Verschlagwortungsmaske ist ein Index vorhanden, zu erkennen an der zugehörigen ID.

Die URL für Statistikdaten lautet: *http://<server>:9200/_stats*.

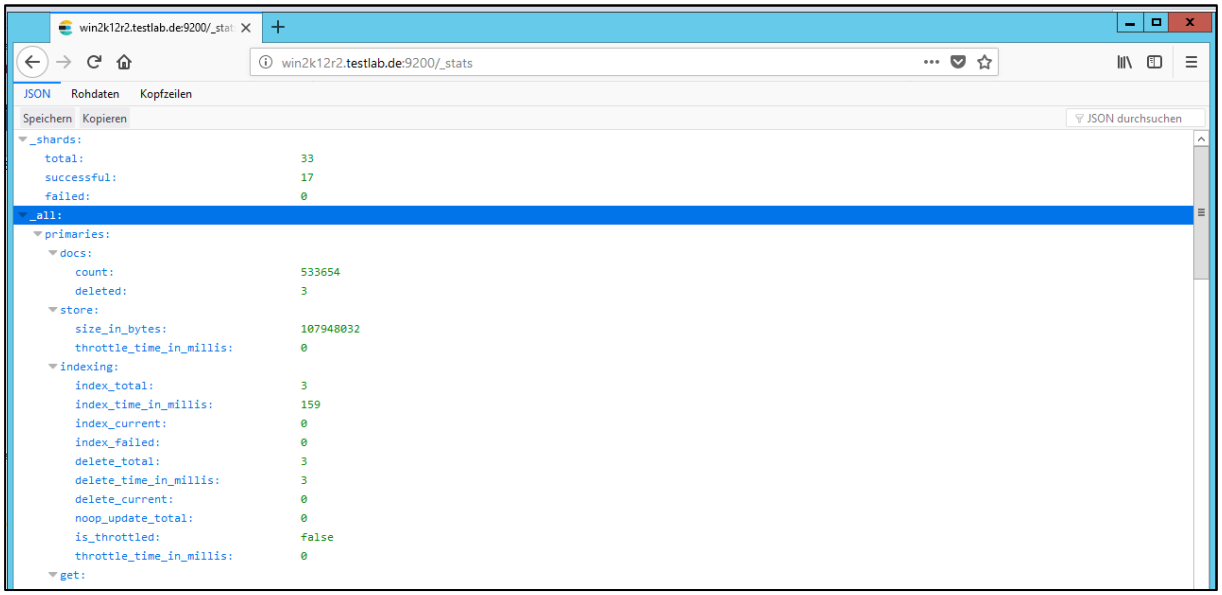


Abb. 44: Ausgabe der statistischen Werte mit dem Firefox-Browser

Hiermit können Sie prüfen, ob ein neues Dokument in den Index aufgenommen wurde. Nach Neuaufnahme eines Dokuments zeigt die Abfrage in der ersten Zeile nach „count" einen um 1 erhöhten Wert an.

In derselben Form sind weitere Abfragen möglich. Diese sind in der Online-Dokumentation der Elasticsearch beschrieben.

4.4.3 Backup-Möglichkeiten und Backup-Strategien

Die Daten der iSearch lassen sich zu 100% aus der Datenbank über den Indexserver wiederherstellen. Die Entscheidung, ob ein Backup des Elasticsearch-Index angelegt wird, ist demnach von folgenden Faktoren abhängig:

- Welche Downtime der iSearch ist für die Nutzer zumutbar?

  Wenn die Downtime möglichst gering sein soll, ist eine Backup-Strategie notwendig, bei der der vorher gesicherte iSearch-Index innerhalb kurzer Zeit in das Produktivsystem gespielt werden kann.

- Wie lange dauert es, den iSearch-Index wiederherzustellen?

  Die Dauer der initialen Re-Indexierung kann als Anhaltspunkt für folgende Re-Indexierungen genommen werden. Es muss entschieden werden, ob diese Dauer mit der Downtime für die Nutzer vereinbar ist.

Zum Erstellen eines Backups gibt es zwei Möglichkeiten. Bei beiden Lösungen müssen Sie beachten, dass die Indizes zwischen unterschiedlichen Indexserver-Versionen nicht unbedingt kompatibel sind. Im Anschluss an das Einspielen des Backups ist es notwendig, die Dokumente, die sich seit dem Erstellen des Backups geändert haben, erneut zu indexieren. Dies lässt sich über den Updater-Prozess der iSearch realisieren.

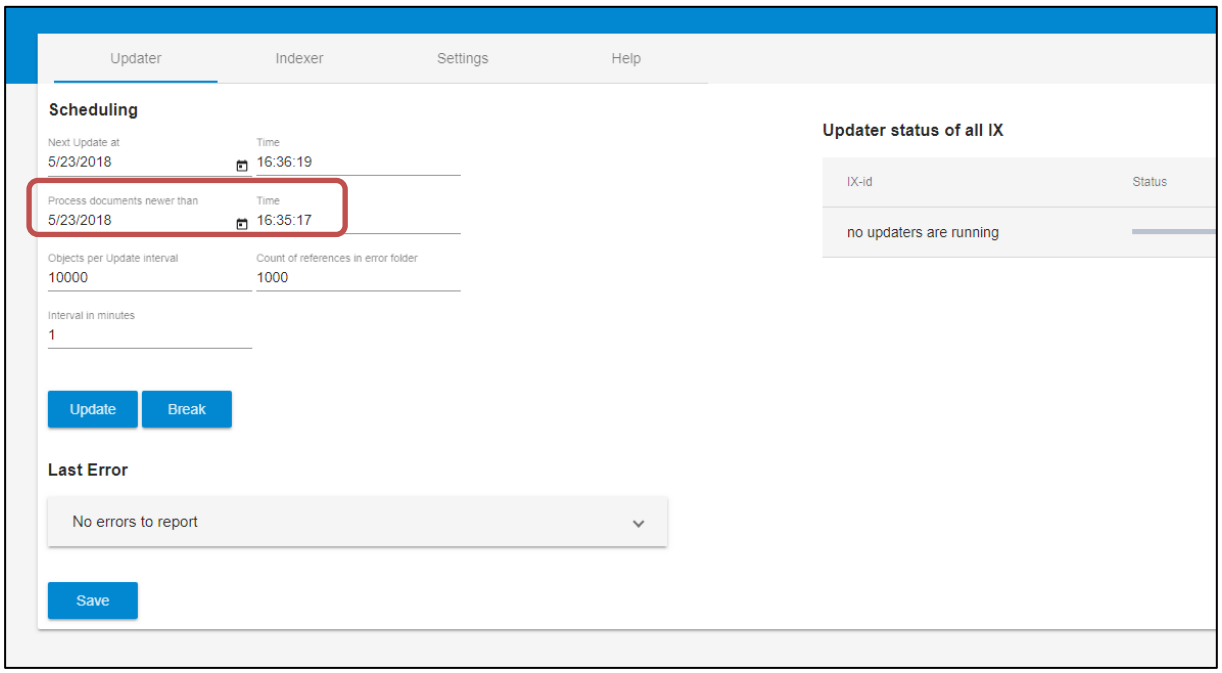


Abb. 45: Setzen des Wertes "Process documents newer than" für den Fall, dass ein Backup eingespielt wird.

Der Zeitpunkt für „Process documents newer than“ wird auf den Zeitpunkt gesetzt, zu dem das Backup erstellt wurde. Demnach ist es sinnvoll, z. B. täglich oder wöchentlich ein Backup anzulegen, damit die Dauer zum Indexieren der geänderten Dokumente nicht zu lange ist.

Zudem ist zu beachten, dass der Index mit dem Namen „searchguard“ benötigt wird, um die Verschlüsselung mittels SSL über das „Search Guard-Plugin“ zu gewährleisten. Existiert dieser Index nicht, so kann sich auch der Indexserver nicht mit der Elasticsearch verbinden. Haben Sie den „searchguard-Index“ aus Versehen gelöscht, können Sie ihn durch erneutes Durchführen des Server-Setups erzeugen. Außerdem ist der „searchguard-Index“ abhängig vom gewählten Tomcat-Administrationsnutzer, so dass der Index bei unterschiedlichen Nutzernamen bzw. Passwörtern nicht einfach kopiert werden kann, sondern durch das Server-Setup erzeugt werden muss.

4.4.3.1 Bestehenden Index kopieren

Den bestehenden Index können Sie aus dem Dateisystem kopieren und zu einem späteren Zeitpunkt wieder einspielen. Dazu kopieren Sie den kompletten Ordner unter *<Installationsverzeichnis>/data/<…><Servername>/index* als Backup. Sollte das Backup benötigt werden, ersetzen Sie den Inhalt des oben genannten Ordners mit dem Backup.

Den iSearch-Dienst sollten Sie vor dem Kopieren stoppen, um Zugriffsfehler zu vermeiden.

4.4.3.2 Snapshot mittels Elasticsearch erstellen

Elasticsearch selbst bietet die Möglichkeit, so genannte Snapshots der bestehenden Indizes anzulegen. Werden regelmäßig Snapshots erzeugt, wird nicht jedes Mal eine komplette Kopie erzeugt, sondern nur die geänderten Dateien gespeichert und so der benötigte Speicherplatz reduziert.

Die Anleitung dazu finden Sie hier:

https://www.elastic.co/guide/en/elasticsearch/reference/5.6/modules-snapshots.html



Beachten Sie: Um einen Snapshot wiederherstellen zu können, muss der Konfigurationsdatei der Elasticsearch (*<Installationsverzeichnis>/config/elastic/<Servername>/elasticsearch.yml*) folgende Zeile hinzugefügt werden:

```
searchguard.enable_snapshot_restore_privilege: true
```

4.5 Tipps und Tricks

4.5.1 ELO iSearch-Konfiguration

Über die Konfigurationsseite des Indexservers, alternativ direkt über den folgenden Link, können Sie die „ELO iSearch Configuration“ aufrufen, die ab ELO 11 zur Verfügung steht:

http://<server>:<Port>:/ix-<Archiv>/manager/esconfig/#/isearchConfig

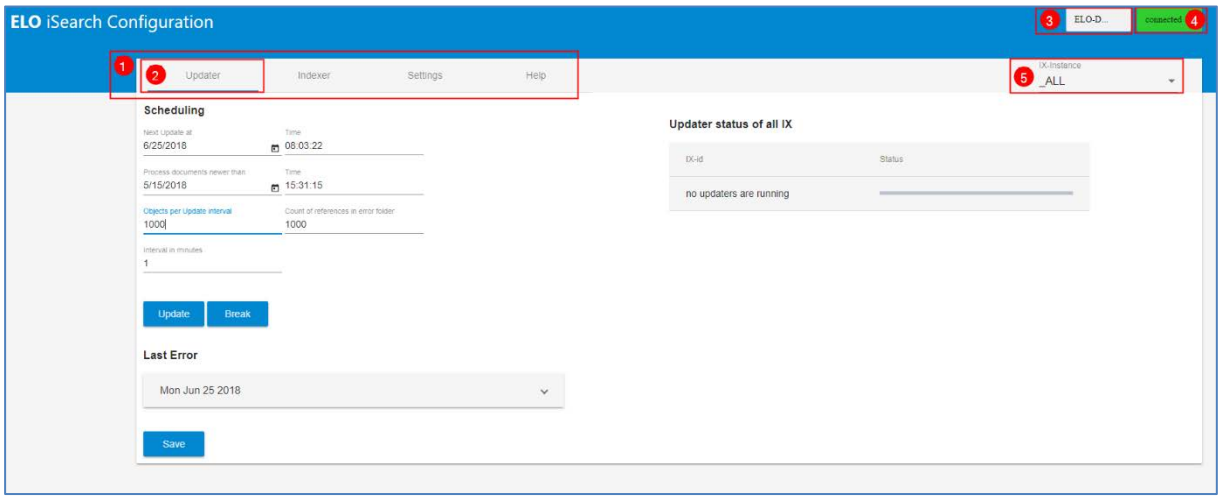


Abb. 46: Übersicht über die ELO iSearch Configuration

In Abb. 46 sehen Sie die ELO iSearch Konfiguration. Sie können über die ELO iSearch Konfiguration die verschiedenen Funktionalitäten der iSearch aufrufen (1), der aktive Tab ist mit einem blauen Unterstrich markiert (2). Oben rechts (3) sehen Sie den Namen der ELOix-Instanz, über die die ELO iSearch Configuration aufgerufen wurde, deren Verbindungsstatus zur iSearch (4) ist „connected“ bzw. „disconnected“.

Ab ELO 11 werden die Einstellungen für die iSearch in der Tabelle eloixopt gespeichert und sind demnach, wie auch die Indexserveroptionen, instanzspezifisch, bzw. für alle Instanzen gespeichert („_ALL“). In der Tabelle haben alle iSearch-Einstellungen das Präfix „esearch“. Über die ELO iSearch Configuration können Sie die Einstellungen ebenso für alle Instanzen („_ALL“) oder instanzspezifisch anzeigen lassen und bearbeiten (5).

4.5.1.1 Updater

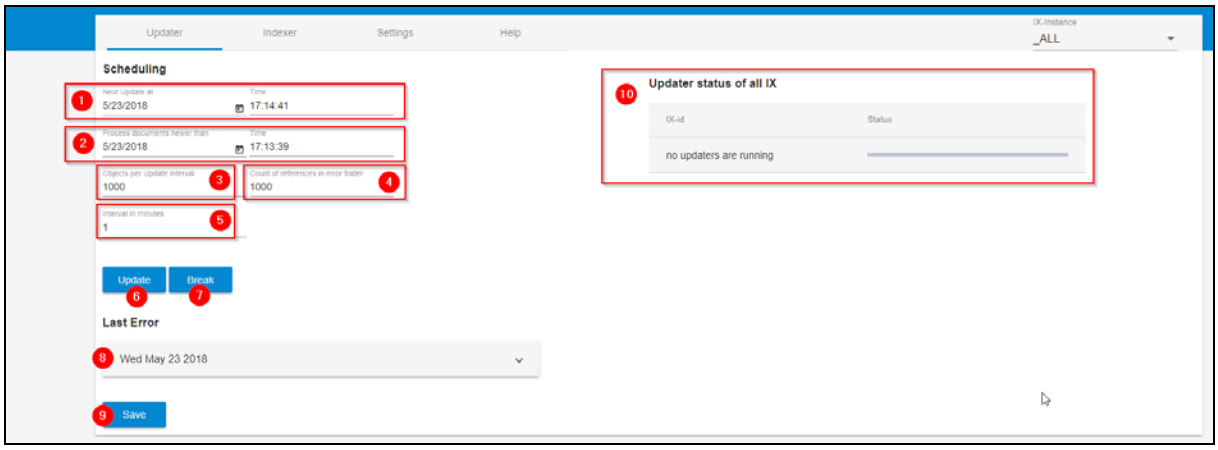


Abb. 47: ELO iSearch Configuration, Tab 'Updater'

In Abb. 47 sehen Sie die Übersicht über den Updater. Diesen benötigen Sie für die regelmäßige Aktualisierung der iSearch. Er sorgt zudem dafür, dass neue und geänderte Objekte an die iSearch gesendet werden. In der obersten Zeile (1) sehen Sie, wann der Updater das nächste Mal startet und in der nächsten Zeile (2), ab welchem Zeitpunkt Dokumente verarbeitet werden. Beide Werte können Sie einstellen. In der Standardkonfiguration läuft der Updater jede Minute (5) und verarbeitet maximal 1000 Objekte (3) in der kleinsten Zeiteinheit (1 Minute). Beachten Sie, dass die Anzahl der ausgewählten Objekte zeitgleich im Speicher des ELOix gehalten wird, den Wert sollten Sie daher nur mit Bedacht anpassen.

Kommt es zu Problemen beim Indexieren eines Dokumentes, wird eine Referenz auf das Dokument im ELO Archiv unter folgendem Pfad angelegt:

Administration // Fulltext Configuration // Update Errors

Dadurch kann ein Administrator leicht kontrollieren, ob es zu Fehlern kam und dafür sorgen, dass die Dokumente erneut an die iSearch gesendet werden, indem z. B. der Verschlagwortungsdialog geöffnet wird. Die maximale Anzahl an Dokumenten in diesem Ordner stellen Sie in (4) ein. Stellen Sie „0“ ein, werden keine Referenzen angelegt.

Nehmen Sie Änderungen vor, müssen Sie diese mit „Save“ speichern (9).

Eine Übersicht über die laufenden Updater-Prozesse ist im Bereich „Updater status of all IX“ (10) dargestellt. Einen Update-Prozess können Sie auch manuell über die Schaltfläche „Update“ (6) starten sowie über die Schaltfläche „Break“ (7) unterbrechen. Beachten Sie hierbei, dass sich das Starten und Unterbrechen immer auf die aktuelle ELOix-Instanz bezieht.

Der letzte Fehler des Updater-Prozesses wird in „Last Error“ (8) angezeigt.

4.5.1.2 Indexer

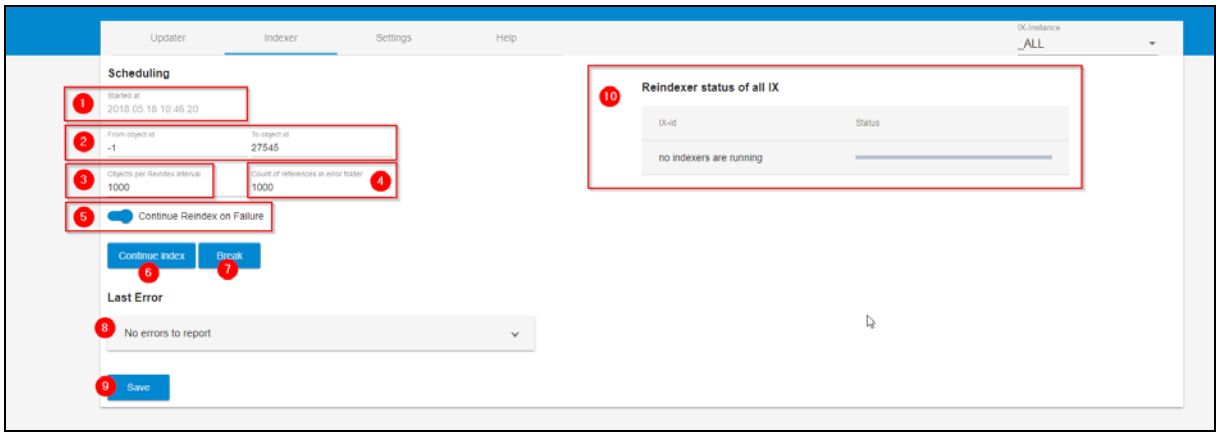


Abb. 48: ELO iSearch Configuration, Tab 'Indexer'

Über den Indexer (Abb. 48) nehmen Sie die initiale Befüllung der iSearch mit den Objekten aus dem ELO Archiv. Dazu werden alle vorhandenen Objekte an die iSearch gesendet. Gestartet wird mit den Objekten mit der höchsten ID, indexiert wird absteigend. Dadurch sind neue Objekte früher auffindbar, sofern der Index sofort produktiv verwendet wird.

In der obersten Zeile (1) ist der Startzeitpunkt des letzten Re-Index-Prozesses zu sehen. Über (2) und (3) wählen Sie die Objekt-IDs aus, die indexiert werden. Möchten Sie einen kompletten Re-Index durchführen, müssen Sie beide Werte auf „-1" setzen. Dies führt auch dazu, dass der bestehende Index gelöscht wird. Möchten Sie jedoch nur bestimmte Objekte erneut einem bestehenden Index hinzufügen, können Sie dies auch über die IDs steuern.

Die Einstellungen für die Objekte pro Indexierungsintervall (3) und Referenzen im Archiv (4) entsprechen denen des Updaters. Die Referenzen werden im Ordner unter folgendem Pfad abgelegt:

Administration // Fulltext Configuration // Re-Index Errors

Dieser Wert ist der Gleiche, sowohl für den Updater als auch für den Indexer.

Die Einstellung „Continue Reindex on Failure" (5) gibt an, ob im Falle eines Fehlers der Re-Index-Prozess unterbrochen oder fortgesetzt werden soll. Möchten Sie ihn fortsetzen, ist ein regelmäßiges Kontrollieren der Log-Ausgaben unablässig.

Der letzte Fehler des Re-Index-Prozesses wird in „Last Error" (8) angezeigt.

Die Einstellungen müssen Sie über die Schaltfläche „Save" (9) speichern.

Ein Re-Index kann gestartet (wenn beide Objekt-IDs auf „-1" gesetzt sind) bzw. fortgesetzt (6) sowie unterbrochen (7) werden. Wie auch beim Updater müssen Sie beachten, dass dies sich immer auf die ELOix-Instanz bezieht, mit der die Config-Seite aufgerufen wurde.

Eine Übersicht über aktive Re-Index-Prozesse ist in (10) dargestellt.

4.5.1.3 Settings

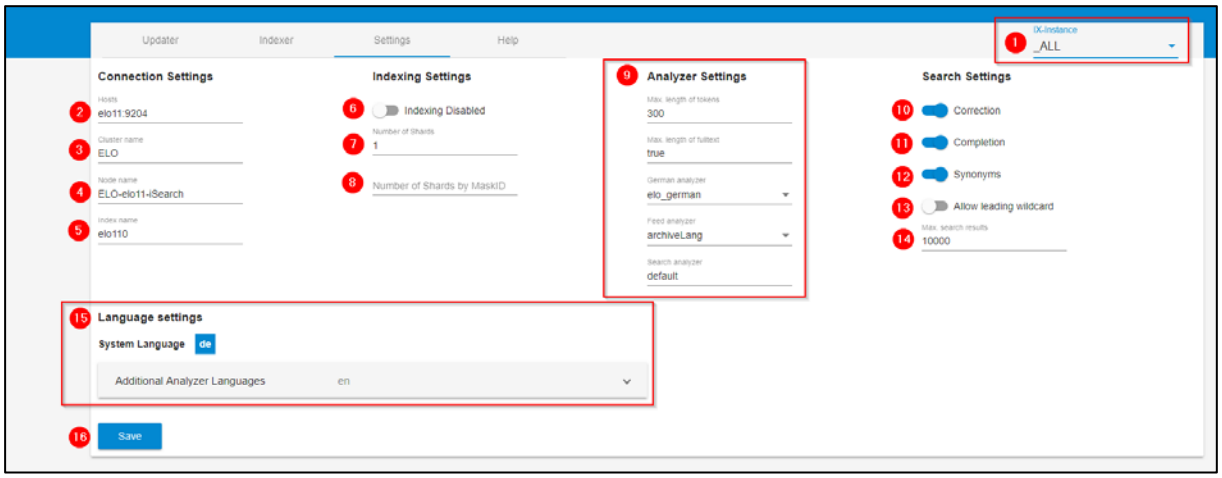


Abb. 49: ELO iSearch Configuration, Tab 'Settings', Einstellungen für '_ALL' vornehmen

Eine Übersicht über die weiteren Einstellungsmöglichkeiten der iSearch ist in Abb. 49 dargestellt. In „IX-Instance“ (1) wählen Sie aus, ob die Einstellungen für alle Instanzen oder instanzspezifisch vorgenommen werden sollen (siehe auch Abb. 50).

Die Details zum Verbindungsaufbau zwischen ELOix und Elasticsearch werden in (2), (3) und (4) angegeben. Der Name des Elasticsearch-Index (5) entspricht im Standard dem Namen des Archivs in Kleinbuchstaben.

Des Weiteren können Sie angeben, ob ein ELOix überhaupt zum Senden von Objekten an die iSearch verwendet werden soll (6). Ist diese Einstellung aktiv, ist die Volltextsuche über die iSearch trotzdem noch aktiv.

Die Anzahl der Shards bzw. Shards pro Maske wird mit (7) und (8) eingestellt. In Kapitel *4.5.5 Festlegung der Shard-Anzahl pro Index* wird näher auf diese Einstellungen eingegangen.

Die Analyzer-Einstellungen (9) beziehen sich darauf, wie Text verarbeitet wird, der an die Elasticsearch gesendet wird. Im Normalfall verwenden Sie die vorhandenen Voreinstellungen.

Die Einstellungen (10) bis (14) beziehen sich auf die Suche. Es ist einstellbar, ob Korrekturvorschläge (10) oder Synonyme (12) zur Suchanfrage zurückgegeben werden und ob die Autovervollständigung (11) bei der Sucheingabe aktiv ist. Des Weiteren können Sie die Suche mit einem Wildcard-Zeichen am Anfang des Suchwortes aktivieren (13). Dies ist standardmäßig deaktiviert, da eine Suche mit führendem Wildcard-Zeichen die Elasticsearch stark auslasten kann. Außerdem können Sie die maximalen Suchergebnisse einstellen (14), die über die ELOix-API-Aufrufe findFirstSords() und findNextSords() zurückgegeben werden.

Die Spracheinstellungen der iSearch können Sie mittels „Language Settings“ (15) bearbeiten. Die System- bzw. Archivsprache wird über das Server Setup eingestellt und kann nicht bearbeitet werden. Je nach verfügbaren Lizenzen können Sie weitere Sprachen auswählen.

Alle Einstellungen müssen Sie über die Schaltfläche „Save“ speichern (16). Beachten Sie, dass die Einstellungen (5), (7), (8), (9) und (15) einen Re-Index benötigen, damit sie für alle Daten gelten und dies nicht zu unterschiedlichen Daten in der iSearch führt.

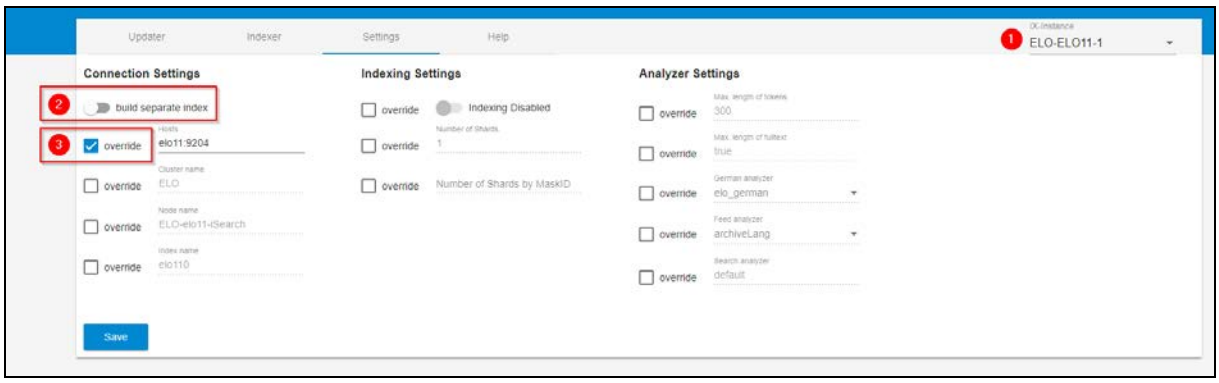


Abb. 50: ELO iSearch Configuration, Tab 'Setting', Einstellungen instanzspezifisch vornehmen

Abb. 50 zeigt beispielhaft die Einstellungsseite für die Instanz ELO-ELO11-1 (1). Alle Einstellungen, bei denen die Override-Checkbox aktiv ist (3), überschreiben die Einstellung, die für alle Instanzen vorgenommen wird.

Die Einstellung „build separate index“ (2) können Sie nur instanzspezifisch vornehmen. Ist diese Einstellung aktiv, wird mit diesem Indexserver ein separater Index aufgebaut – allerdings nur dann, wenn sich entweder die Verbindungseinstellungen, von denen für die anderen Instanzen unterscheiden oder der Indexname unterschiedlich ist. Im Kapitel *4.2.4 Volltextindexaufbau im Hintergrund* wird dies näher erläutert.

4.5.2 Dienstekonfiguration der Elasticsearch

Die Konfiguration des Elasticsearch-Dienstes erfolgt über mehrere Dateien und Verzeichnisse. Siehe beispielhaft im folgenden Screenshot.

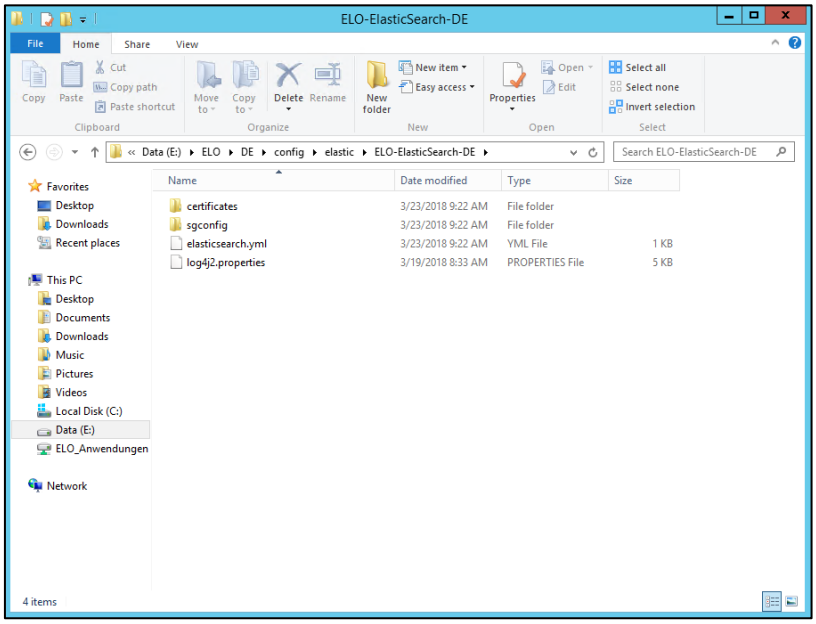


Abb. 51: Dateien und Verzeichnisse zur Konfiguration des Elasticsearch-Dienstes

Im Einzelnen haben die Dateien und Verzeichnisse folgende Inhalte:

- *elasticsearch.yml*: Zentrale Konfigurationsdatei der Elasticsearch. Detaillierte Informationen lesen Sie unter folgendem Link: *https://www.elastic.co/guide/en/elasticsearch/reference/current/important-settings.html*
- log4j.properties: Logging-Einstellungen
- sgconfig: Verzeichnis mit Konfigurationen des „Search Guard Plugins“
- certificates: Key- und Truststore für die SSL-Kommunikation mit dem ELO Indexserver

4.5.3 SSL-Kommunikation

Ab der ELO ECM Suite 10.2 kommunizieren die ELO Indexserver verschlüsselt mit der Elasticsearch. Dazu werden vom ELO Server Setup bei der Installation „selfsigned Zertifikate“ und entsprechende Key- und Truststores angelegt. Die Key- und Truststores werden unterhalb der jeweiligen config-Verzeichnisse (Indexserver und Elasticsearch) angelegt und können mit Java-Standardmitteln (z. B. „keytool.exe“) für die Prüfung der Zertifikate geöffnet werden. Als Passwort wird das Tomcat-Administrator-Passwort verwendet.

4.5.4 Indexserver-Plugins

Ab ELO 11 ist es möglich, den Indexserver durch sogenannte OSGi-Plugins zu erweitern. Beispielsweise befindet sich die Elasticsearch-Anbindung selbst in einem solchen Plugin. Mit dieser Aufteilung ist es zukünftig möglich, mit einer älteren Indexserver-Version auf einen neueren Elasticsearch-Index zuzugreifen. Für jeden Indexserver wird ein Plugin-Verzeichnis konfiguriert (Pfad siehe Abb. 52).

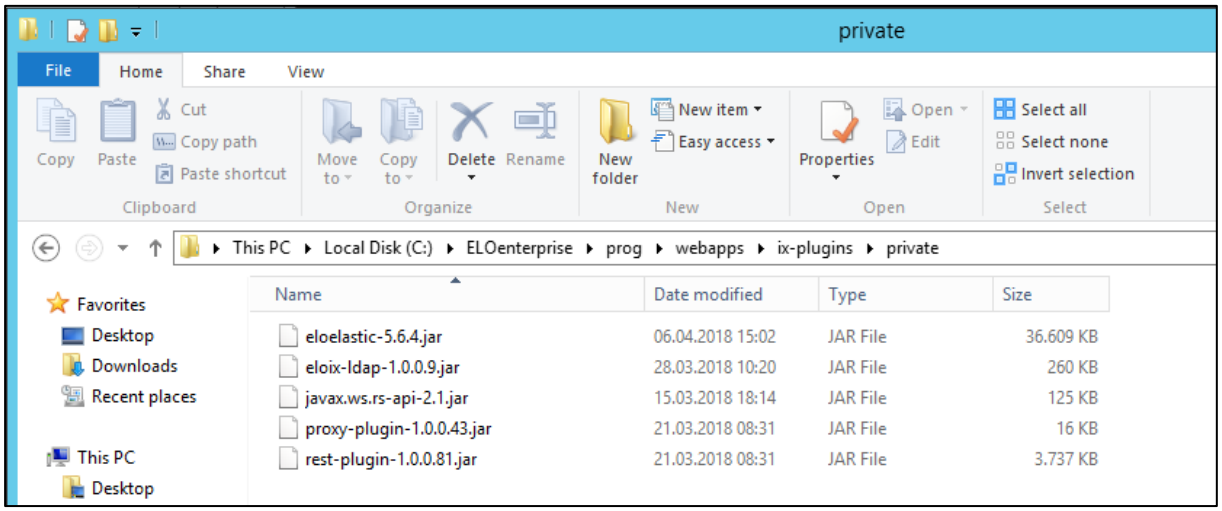


Abb. 52: Pfad des Indexserver-Plugin-Verzeichnisses

4.5.5 Festlegung der Shard-Anzahl pro Index

Ab ELO 11 wird pro Maske ein Index angelegt. In der Standardeinstellung besteht jeder Index aus einem Shard. Durch die Aufteilung in einen Index pro Maske wird auch im Cluster schon ein Performance-Gewinn erreicht, da die Indizes auf mehrere Nodes aufgeteilt werden.

Zusätzlich ist es sinnvoll, besonders große Indizes in mehrere Shards aufzuteilen. Der Vorteil liegt darin, dass Suchanfragen dann parallel auf allen Shards ausgeführt werden. Allerdings werden maximal so viele Suchanfragen innerhalb eines Nodes parallel ausgeführt wie Prozessoren vorhanden sind.

Die Anzahl an Shards richtet sich demnach nach den enthaltenen Daten. Elasticsearch empfiehlt eine Shard-Größe von 20 GB – 40 GB.[1] Zu kleine Shards führen zu einer Überlastung in der Kommunikation zwischen den Shards, da die Daten zwischen den Shards synchronisiert werden müssen. Masken, die sehr viele Sords mit Volltext enthalten, sollten demnach möglichst in mehrere Shards aufgeteilt werden.

Die Anzahl an Shards pro Maske kann über die ELO iSearch Configuration vorgenommen werden, erfordert allerdings einen Volltextindexaufbau (Re-Index) aller Daten.

[1] https://www.elastic.co/blog/how-many-shards-should-i-have-in-my-elasticsearch-cluster

4.5.6 Festlegung der Replica-Anzahl pro Index

Replicas sind gewissermaßen Kopien eines Index. Es kann jeder beliebige Wert gewählt werden, auch 0. Für jeden Shard wird ein Primary-Shard erzeugt, sowie im Cluster die angegebene Anzahl an Replica-Shards. Primary-Shards und Replica-Shards werden immer auf unterschiedlichen Nodes angelegt, um im Falle eines Ausfalls auch weiterhin die Daten bereitstellen zu können. Fällt ein Node aus, werden die Primary-Shards, die bisher auf diesem Node allokiert waren, auf einem anderen Node allokiert. Die Replica-Shards werden auf andere Nodes verteilt, um weiterhin die bestmögliche Ausfallsicherheit und Performance zu gewährleisten.

Der weitere Vorteil von Replica-Shards ist die Parallelisierung von Suchanfragen: Die Elasticsearch verteilt Suchanfragen auf unterschiedliche Primary-Shards und Replica-Shards, um die Last der Suchanfragen zu verteilen.

Der Health-Zustand eines Indizes ist abhängig davon, welche Shards verfügbar sind: Ist der Primary-Shard vorhanden, ist der Zustand „gelb“, sind zusätzlich alle Replica-Shards verfügbar, ist der Zustand „grün“.

Die Menge der Replica-Shards sollten Sie an die Anzahl der Nodes im Cluster anpassen, so dass sich die Shards sinnvoll auf mehrere Nodes verteilen können, um Ausfallsicherheit zu gewährleisten. Andererseits benötigt jede Replica genauso viel Speicherplatz wie die Primary-Shard.

Die Standardeinstellung für die Anzahl an Replicas ist 1. So wird in einem Cluster von mehr als einer Instanz jeder Shard einmal repliziert, ist also insgesamt zweimal vorhanden. Eine Änderung an der Anzahl der Replicas erfordert keine Re-Indexierung.

Um eine Re-Indexierung in einem Cluster zu beschleunigen, können Sie für die Dauer der Re-Indexierung die Anzahl an Replicas auf 0 setzen und nach dem Beenden erhöhen. Ist die Anzahl an Replicas > 0, werden die Daten sowohl im Primary-Shard als auch in jedem Replica-Shard zeitgleich indexiert. Das bedeutet, dass die Analyse und Umwandlung in Elasticsearch-Daten mehrmals parallel stattfinden. Wird die Anzahl an Replicas auf 0 gesetzt, werden die Daten während der Re-Indexierung nur in einem Node verarbeitet. Sobald die Anzahl an Replicas erhöht wird, werden die bereits indexierten Primary-Shards auf die anderen Nodes kopiert, was in der Regel deutlich schneller geht. Natürlich ist dadurch während der Re-Indexierung keine Ausfallsicherheit gewährleistet.

Die Einstellung ist bezogen auf einen Index und kann z. B. mit dem curl-Tool (*https://curl.haxx.se/*) exemplarisch folgendermaßen für die Maske mit der ID 42 des Archivs „archiv“ auf 2 geändert werden:

```
curl -X PUT "localhost:9200/archiv¶42/_settings" -H 'Content-Type:
application/json' -d'
{
    "index" : {
        "number_of_replicas" : 2
    }
}
'
```

Möchten Sie die Einstellung für alle Indizes ändern, können Sie die Angabe des Index weglassen und die URL `localhost:9200/_settings` verwenden.

4.5.7 Tokenisierung

Die Option „Tokenisierung deaktivieren“ einer Indexfeldvorlage schaltet die Suchregeln und -muster für dieses Feld aus. Es erscheinen dann in der Liste des Filters die exakten Werte der Verschlagwortung. Die Suche arbeitet für dieses Feld dann ebenfalls in einem exakten Modus. Teilbegriffe werden in diesem Fall nicht berücksichtigt.

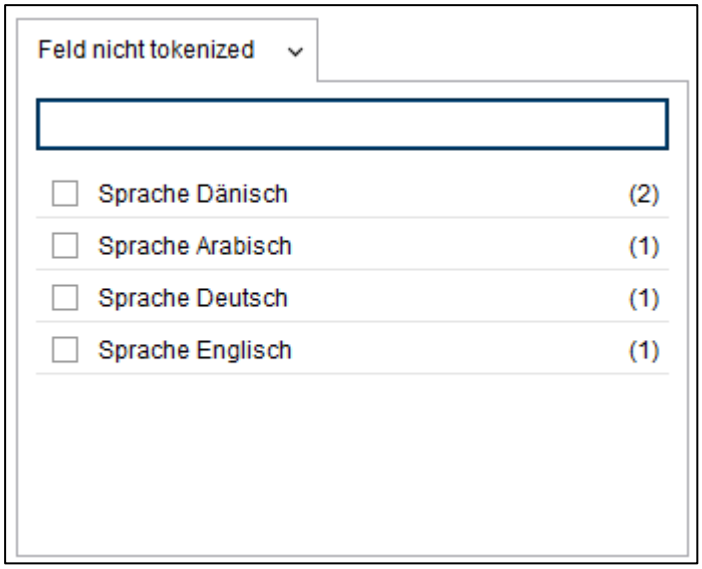


Abb. 53: Indexfelder nicht in Token aufgeteilt

Bei aktiven Suchregeln und -mustern erscheinen in der Liste des Filters nur einzelne Wörter, sowohl die Originale als auch die erzeugten Stammformen. Die Suchmöglichkeiten in diesem Filter entsprechen dann denen des Eingabefeldes der Suche.

4.5.8 Einbinden eigener Thesauri

Die Wörterbücher für den Thesaurus der ELO iSearch können Sie für ein Archiv individuell anpassen. Die benötigten Dateien liegen im Archiv unter folgendem Pfad:

Administration // Fulltext Configuration // Thesaurus

In diesem Pfad befindet sich für jede Sprache ein Unterverzeichnis mit dem Länderkürzel, in dem sich jeweils drei Dokumente befinden.

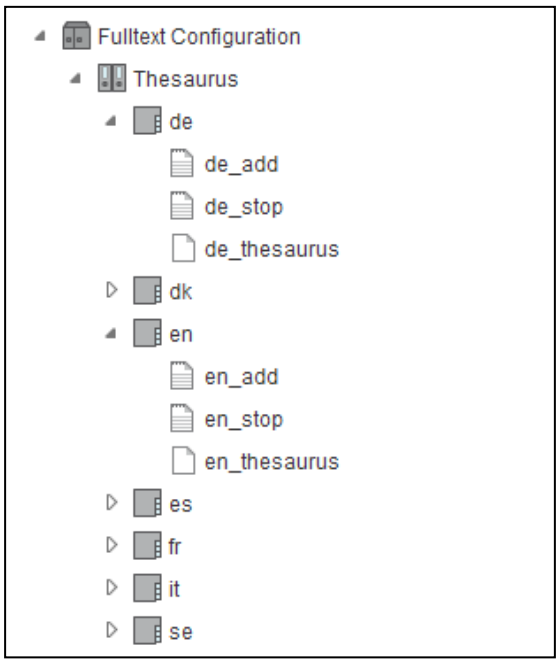


Abb. 54: Struktur der Thesaurus-Wörterbücher

<Sprache>_thesaurus: Der von ELO ausgelieferte Thesaurus für diese Sprache. Diese Datei sollten Sie nicht ändern, damit sie bei einem Update gegen eine neuere Version ausgetauscht werden kann.

<Sprache>_add: In diesem Dokument können Sie zusätzliche Begriffe eintragen, welche als Synonyme in den Thesaurus aufgenommen werden sollen.

<Sprache>_stop: Über dieses Dokument können Sie einzelne Begriffe aus dem Thesaurus ausschließen, wenn diese Ihnen für Ihren Einsatz unpassend erscheinen.

Beachten Sie: Nach einer Änderung der Dokumente muss der ELO Indexserver neu gestartet werden, damit der Thesaurus anhand der geänderten Dateien neu aufgebaut werden kann.

4.5.9 Cluster-Betrieb

Im ELO SupportWeb wird ein Dokument zum Cluster-Betrieb von ELO veröffentlicht, das weitere Details zum Betrieb der iSearch enthält.

4.6 Powersuche

Die Powersuche der ELO iSearch bietet die Möglichkeit, komplexe Anfragen an die iSearch zu stellen und direkt auf deren Syntax zuzugreifen.

Die Powersuche der ELO iSearch erlaubt es Benutzern, direkt auf die zugrundeliegende Elasticsearch zuzugreifen. Dadurch bieten sich zwei neue Möglichkeiten: Zum einen kann auf Felder, die nicht über einen Filter auswählbar sind, direkt zugegriffen werden. Zum anderen können über die von Elasticsearch angebotene Syntax sehr komplexe Anfragen gestellt werden.

Der direkte Zugriff auf die internen Strukturen ist gleichzeitig auch der Nachteil der Powersuche, denn die Schreibweise ist sehr technisch und erfordert Wissen über die verwendeten internen Datenstrukturen. Dieser interne Aufbau kann sich zudem in kommenden ELO Versionen ändern, sodass vorhandene Suchabfragen nicht mehr richtig oder gar nicht mehr funktionieren.

Achtung: Die Powersuche sollte nur in besonderen, einzelnen Fällen benutzt werden, wenn die normalen Suchmöglichkeiten nicht ausreichen (z. B. bei administrativen Spezialanfragen).

4.6.1 Anwendung der Powersuche

Die Powersuche kann sowohl über den ELO Java Client als auch den ELO Web Client verwendet werden. Die Suchanfrage wird über die Suchzeile eingegeben und mit einem „=“ ausgelöst, wie im folgenden Beispiel zu sehen ist.

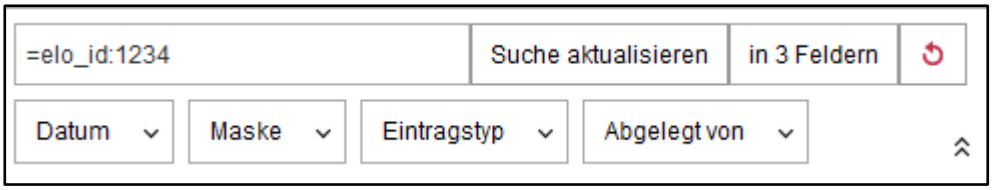


Abb. 55: Beispiel für eine Powersuche im ELO Java Client

Mit der Powersuche werden Anfragen über den ELO Indexserver direkt an die Elasticsearch geschickt. Der ELO Indexserver fügt eine Berechtigungsüberprüfung hinzu, sodass ein Benutzer trotz Powersuche immer nur Dokumente angezeigt bekommt, für die er eine Berechtigung hat.

Die Syntax der Powersuche entspricht der Syntax der „QueryStringQuery“, wie sie von Elasticsearch unter diesem Link dokumentiert wird:

https://www.elastic.co/guide/en/elasticsearch/reference/5.6/query-dsl-query-string-query.html#query-string-syntax

Für den Benutzer ist vor allem wichtig, dass Sonderzeichen nun von ihm selbst escaped werden müssen. Dies ist notwendig, wenn ein Sonderzeichen wie ein normaler Buchstabe interpretiert werden soll und nicht in seiner besonderen Bedeutung, welche die Suche beeinflusst. Folgende Sonderzeichen haben eine solche besondere Bedeutung in der „QueryStringQuery"-Syntax und müssen, wenn diese Bedeutung nicht gewünscht ist, escaped werden:

Zeichen	Bedeutung
+<wort>	<wort> muss im Dokument vorkommen
-<wort>	<wort> darf im Dokument nicht vorkommen (nur wenn davor ein Leerzeichen steht)
<wort1> && <wort2>	Beide Worte müssen im Dokument vorkommen
<wort1> || <wort2>	Mindestens eines der beiden Worte muss im Dokument vorkommen
> x < x	Größer/ kleiner als …-Suche (funktioniert mit Zahlen oder mit Worten)
>= x <= x	Größer gleich/kleiner gleich als …-Suche (funktioniert sowohl mit Zahlen als auch mit Worten)
=	Allein keine Bedeutung, am Anfang der Zeile: Powersuche
{x TO y}	Bereichssuche (ohne x und y)
[x TO y]	Bereichssuche (inklusive x und y)
<wort>^	Gewichtung des Wortes
„<wort> <wort>"	Phrasensuche
<wort>~	Fuzziness (unscharfe Suche)
*	Wildcard, beliebig viele Zeichen
?	Wildcard, exakte Anzahl an zu ersetzenden Zeichen
\	Escaped-Zeichen
/	Keine Bedeutung
<feldname>:<wort>	Suche in <feldname> nach <wort>

4.6.2 Feldnamen

Einzelne Felder der Elasticsearch können direkt abgefragt werden, sofern der Feldname bekannt ist. Eine Übersicht über die Feldnamen wird von ELO nicht bereitgestellt. Ein Administrator kann sie jedoch über die REST-Schnittstelle der Elasticsearch abfragen:

```
http://<host>:<port>/<Archivname in Kleinbuchstaben>/_mapping?pretty
```

Weitere Informationen zur Interpretation der Abfrage bietet Elasticsearch unter diesem Link: *https://www.elastic.co/guide/en/elasticsearch/reference/5.6/index.html*



Beachten Sie: ELO behält es sich vor, die Feldnamen zwischen den Versionen zu ändern. Daher ist es möglich, dass eine Anfrage mit einer Version funktioniert, mit der nächsten Version jedoch nicht mehr.

4.6.3 Beispiele für Abfragen

Oft ist es sinnvoll herauszufinden, ob ein bestimmtes Dokument über die iSearch auffindbar ist. Dies kann über folgende Suchanfrage herausgefunden werden:

```
=elo_id:<Objektid>
```

Voraussetzung für die Suche ist, dass man als Benutzer mit dem Recht „Alle Einträge sehen, Berechtigungen ignorieren" sucht. Falls dies zu keinem Ergebnis führt, ist das Dokument bisher nicht in die iSearch aufgenommen worden.

Die Suche nach *ELO[123* führt zu einem Syntaxfehler, da die eckige Klammer ein Sonderzeichen ist, das auch mit der Standardsuche für eine Bereichssuche genutzt werden kann. Um trotzdem nach diesem Zeichen zu suchen, kann die Powersuche verwendet werden, indem folgende Abfrage gestellt wird:

```
=ELO\[123
```

4.7 ELO Analytics

ELO Analytics verwendet die iSearch für die grafischen Darstellungen. Die Daten werden in einem eigenen Index „kibana" der iSearch gespeichert. Weitere Details enthält das Dokument „ELO Analytics – Technische Dokumentation".

4.7.1 ELO Analytics-Komponentenübersicht

Wichtig für das Verständnis ist die Interaktion zwischen iSearch, ELO Indexserver, ELO Analytics und ELOwf. Das folgende Diagramm zeigt eine Übersicht zu ELO Analytics und die Kommunikation der einzelnen Komponenten.

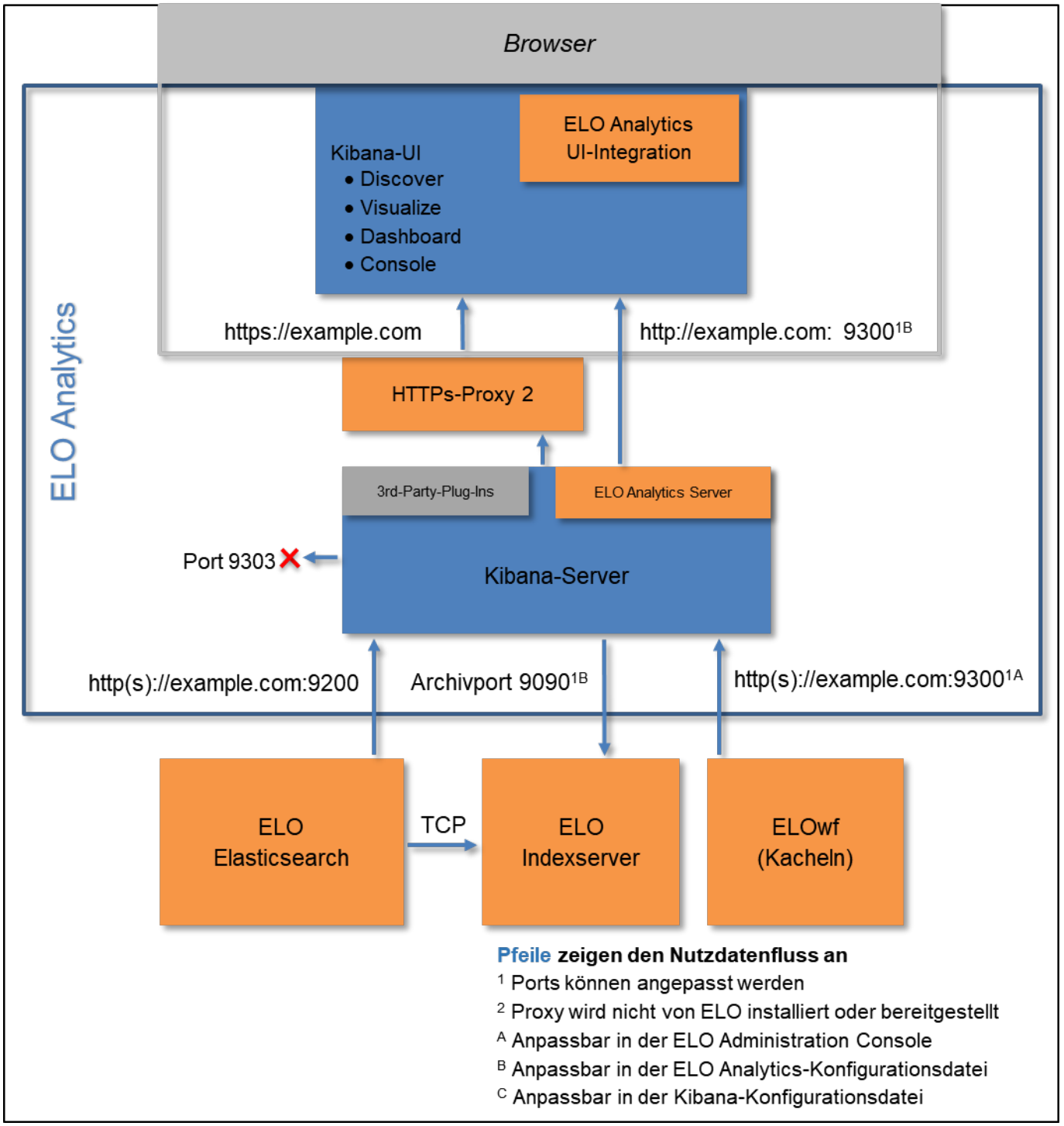


Abb. 56: ELO Analytics Komponentendiagramm

5 Lizenzbestimmungen bei der Verteilung von Serverprozessen

ELOprofessional-Systeme sind nur bedingt clusterfähig, sie erlauben ausschließlich den Hot-Standby-Betrieb der einzelnen Komponenten. ELOenterprise-Systeme sind dagegen voll clusterfähig.

Bei ELOprofessional dürfen die Serverprozesse nur innerhalb desselben Hosts, also auf einem Serversystem, verteilt werden. Bei ELOenterprise hingegen lassen sich die Prozesse auf mehrere Hosts (VMs oder physische Server) verteilen.

Funktionsbereiche	ELOenterprise	ELOprofessional
Betriebssystem	Plattformunabhängig[2]	Windows[1]
Architektur	SOA-konforme modulare Client-Server Architektur	SOA-konforme modulare Client-Server Architektur
Max. Anzahl Archive	unbegrenzt	20 pro Archivserver
Max. Anzahl Anwender	unbegrenzt	1000
Max. Anzahl Dokumente	2 Mrd. pro Archiv	100 Mio
Max. Archivebenen	256	256
Datenbank	PostgreSQL / MS SQL / Oracle / IBM DB2 [3]	PostgreSQL / MS SQL / Oracle / IBM DB2 [3]
Mandantenfähig	✓	•
Portalanbindung	✓	✓
Statuslog für zentrale Monitor-Systeme	✓	✓
Verteilbarkeit von Serverprozessen	✓	Single Host
Clusterfähigkeit	✓	Hot-Standby

Abb. 57: Auszug ELO Produktvergleich: Unterschiede der Systeme ELOenterprise/ELOprofessional

Im Rahmen von ELOprofessional können somit die ELO Serverkomponenten (z. B. ELO Automation Service (ELOas)) auf mehreren Tomcat-Servern verteilt werden, solange sich diese auf einem Serversystem (VM oder physischer Server) befinden. Bei ELOenterprise ist es hingegen erlaubt, die Tomcat-Server auf mehreren Serversystemen (VMs oder physischen Servern) zu installieren.

6 Glossar

Escapen	Das Maskierungszeichen (Escape-Zeichen) ist ein bestimmtes Zeichen, das verhindert, dass das nachfolgende Zeichen vom ausführenden Programm als Funktionszeichen erkannt wird. Das Escapen ist also dann notwendig, wenn ein Sonderzeichen wie ein normaler Buchstabe interpretiert werden soll und nicht in seiner besonderen Bedeutung.
Garbage Collection (GC)	Die **Garbage Collection**, kurz **GC** (englisch: „Müllabfuhr", auch Automatische Speicherbereinigung oder Freispeichersammlung genannt) bezeichnet in der Software- und Informationstechnik eine automatische Speicherverwaltung, die den Speicherbedarf eines Computerprogramms minimiert. Dabei wird zur Laufzeit versucht, nicht länger benötigte Speicherbereiche automatisch zu identifizieren, um diese dann freizugeben. (Quelle: https://de.wikipedia.org/wiki/Garbage_Collection, Stand, 25.03.2019)
Lucene	Apache **Lucene** ist eine hoch performante, frei verfügbare Programmbibliothek, die sich besonders gut zur Volltextsuche eignet. Lucene kann jede Art von Dokumenten indizieren. Dies können Texte sein, aber auch Zahlen, Datumswerte oder GPS-Positionen. Lucene ist die unter der Such- und Analytik-Engine Elasticsearch liegende Struktur zur Datenspeicherung.
NODE / NODES	**NODE** – engl. node bedeutet Knoten – **NODES** – Plural. Ein Elasticsearch-Server entspricht einem Node.
OCR	Optische Zeichenerkennung (englische Abkürzung **OCR** von „optical character recognition") bezeichnet die automatische Texterkennung innerhalb von Bildern. Gescannte Dokumente werden durch die OCR in eine Textform umgewandelt.
OSGi (OSGi-Plugins	Die **OSGi Alliance** (früher *Open Services Gateway initiative*) spezifiziert eine hardwareunabhängige dynamische Softwareplattform, die es erleichtert, Anwendungen und ihre Dienste per Komponentenmodell („Bundle"/„Service") zu modularisieren und zu verwalten („Service Registry"). Die OSGi-Plattform setzt eine Java Virtual Machine (JVM) voraus und bietet darauf aufbauend das OSGi-Framework. (Quelle: https://de.wikipedia.org/wiki/OSGi, Stand, 25.03.2019)

REST	**Representational State Transfer** (abgekürzt **REST**, seltener auch **ReST**) bezeichnet ein Programmierparadigma für verteilte Systeme, insbesondere für Webservices. REST ist eine Abstraktion der Struktur und des Verhaltens des World Wide Web. REST hat das Ziel, einen Architekturstil zu schaffen, der die Anforderungen des modernen Web besser darstellt. Dabei unterscheidet sich REST vor allem in der Forderung nach einer einheitlichen Schnittstelle. Der Zweck von REST liegt schwerpunktmäßig auf der Maschine-zu-Maschine-Kommunikation. (Quelle: https://de.wikipedia.org/wiki/Representational_State_Transfer, Stand, 25.03.2019)
Replikation	**Replikation** bezeichnet die mehrfache Speicherung derselben Daten. In der Elasticsearch kann man mit dieser Backup-Funktion einstellen, wie viele Replikate man von einem Shard haben möchte. Diese werden dann intelligent auf die Cluster-Nodes aufgeteilt. Somit wird die Suche auf mehrere Nodes verteilt und beim Ausfall eines Nodes sind Replikationen der Daten vorhanden, so dass die Suche ohne Datenverlust weiterhin verfügbar ist.
Shard	Die Daten eines Elasticsearch-Suchindex werden in **Shards** aufgesplittet (wörtlich übersetzt „Scherbe“).
Sprach-Analyzer	Die Elasticsearch wird mit Analyzern der gängigsten Sprachen ausgeliefert. Diese **Sprach-Analyzer** führen im Wesentlichen vier Funktionen aus: Sie teilen den Text in einzelne Wörter auf, setzen die Token in Kleinbuchstaben um, entfernen gängige Stoppwörter und leiten Token zu ihrer Stammform ab.
Suchindex	Elasticsearch speichert alle Daten in einem **Suchindex**. Dieser Index wiederum ist aufgesplittet in Typen und Dokumente mit Eigenschaften.

SWAP-Modus / Swapping	In der Informatik beschreibt **Swapping** (engl. für Umlagerung; von englisch to swap, ‚austauschen') innerhalb der Speicherhierarchie das Auslagern von Daten einer höheren in eine tiefere Ebene oder das entgegengesetzte Laden. Durch Swapping sollen die Vorteile höherer Ebenen, also höhere Geschwindigkeit, und tieferer Ebenen, nämlich höhere Kapazität und kostengünstigerer Speicher, quasi gleichzeitig nutzbar gemacht werden. (Quelle: https://de.wikipedia.org/wiki/Swapping, Stand, 25.03.2019)
Token	Beim Erstellen oder Aktualisieren eines Datensatzes zerlegt die Elasticsearch den durchsuchbaren Text des Datensatzes in einzelne Begriffe, die sogenannten **Tokens**. Anschließend werden die Tokens im Suchindex gespeichert. Eine Suchanfrage wird wiederum in Tokens zerlegt. Nach diesen wird dann im Suchindex gesucht. Ein Token entspricht nicht immer einem grammatikalisch korrekten Wort, sondern ist eine Zerlegung und Vereinfachung von Worten, so dass auch eine Suche nach „Auto" ein Dokument findet, das nur das Wort „Autos" enthält.
Tomcat	Apache **Tomcat** ist ein Open-Source-Webserver und Webcontainer, der die Spezifikation für Java Servlets und JavaServer Pages (JSP) implementiert und es damit erlaubt, in Java geschriebene Web-Anwendungen auf Servlet- beziehungsweise JSP-Basis auszuführen.
YML	**YML** ist die Dateierweiterung der vereinfachten Auszeichnungs-sprache YAML. YAML wird zur Datenserialisierung verwendet und lehnt sich an XML und die Datenstrukturen in den Sprachen Perl, Python und C an.

Revisionsgeschichte dieses Dokuments			
Version	Datum	Bearbeiter	Änderungen
1	12.07.2017	bh	Erste Version für ELO 10
2	05.07.2018	bh	Version für ELO 11
3	28.03.2019	bh	Version für ELO 12
4	27.09.2019	fs, bh	Hinweis zur Trefferliste (3.6)